Genes, Environment, and Epigenetics in Neural Tube Defects

by

Deidre Renee Krupp

Department of Molecular Genetics and Microbiology
Duke University

Date:_____

Approved:

_____
Simon Gregory, Supervisor

_____
Beth Sullivan

_____
Debra Silver

_____
Gregory Crawford

_____
Micah Luftig

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Molecular Genetics and Microbiology
in the Graduate School of Duke University

2014

ABSTRACT

Genes, Environment, and Epigenetics in Neural Tube Defects

by

Deidre Renee Krupp

Department of Molecular Genetics and Microbiology
Duke University

Date:_____

Approved:

_____
Simon Gregory, Supervisor

_____
Beth Sullivan

_____
Debra Silver

_____
Gregory Crawford

_____
Micah Luftig

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Molecular Genetics and Microbiology
in the Graduate School of Duke University

2014

Abstract

Neural tube defects (NTDs) are a common class of human birth defects with a complex, multifactorial etiology. Although many contributing factors have been identified, an estimated 60% of human population risk remains unexplained. A portion of that risk is likely attributable to gene-gene and gene-environment interactions which have yet to be fully elucidated. In one project, we used whole-exome sequencing to identify candidate genetic factors in a multiplex anencephaly family, revealing an aggregation of rare and common variants in planar cell polarity genes among the affecteds. In the second project, we profiled the methylomes of a pair of monozygotic twins discordant for anencephaly and identified several differentially methylated sites which could contribute to NTD risk, particularly the *mir-886* locus. Finally, we performed whole-exome and whole-methylome sequencing of mouse strains with differential susceptibility to fumonisin-induced NTDs, in combination with a human SNP association study. We identified epigenetic changes and variant associations which implicate Wnt and Hippo signaling genes as modifiers of the metabolic impacts of fumonisin exposure. These findings underscore the complexity of NTD pathogenesis and highlight the need to elucidate gene-gene and gene-environment interactions contributing to NTD etiology.

Contents

List of Tables

List of Figures

Acknowledgements

The research presented in this dissertation could not have happened, and did not happen, in a vacuum. I would first like to thank the members of the Duke NTD team, both past and present, whose help and support have been invaluable. Karen Soldano (research analyst) and Melanie Garrett (statistical analyst) contributed the human genotype data generation and analysis in Chapter 4. I would also like to thank Allison Ashley-Koch, who as co-PI of the NTD team has been an unofficial mentor to me.

A number of the experiments presented here are products of a collaboration led by Janee Gelineau-van Waes of Creighton University. Researchers from the Centro de Investigaciones en Nutrición y Salud and Diagnóstico Molecular in Guatemala City, Guatemala, under Olga Torres; and from the USDA-ARS Toxicology and Mycotoxin Research Unit in Athens, GA, under Ronald Riley also contributed to this project. Each group brought a complementary specialty to the table, creating a synthesis definitively greater than the sum of its parts. I have greatly enjoyed working with all the researchers on this project; to me, this has been an exemplar of what can be accomplished when researchers with different specialties focus on one issue, however complex.

I would also like to thank all of my co-workers at the Duke Center for Human Genetics for the benefits of their experience and moral support throughout my graduate studies. I particularly appreciate the assistance and expertise of Julie Rochelle in

instruments, protocols, and everyday operation of the lab; Beth Rusnak in maintenance of the analysis computers and the patient installation of every software package I ever wanted to test; and Karen Abramson in all things relating to Illumina. Additionally, the members of the Gregory lab have been invaluable for their practical and moral support, and in providing a diverse array of baked goods over the years.

I am especially grateful to the participants of our NTD study, without whom this research would have been impossible. I hope the results of my studies will help resolve outstanding questions in the field of NTDs, and eventually lead to new therapeutic interventions from which everyone can benefit. Part of this research was funded by the generosity of participants. Financial support also came from the David H. Murdock Research Institute in Kannapolis, NC and NIH grants HD067971 and NS039818.

I would like to thank my dissertation committee members, Beth Sullivan, Debby Silver, Greg Crawford, and Micah Luftig for all of their constructive feedback and support throughout my graduate studies.

Finally, I am grateful to Simon Gregory, my graduate advisor and mentor, who has been supportive of all my endeavors. His mentorship has provided many opportunities for me to broaden my horizons and hone my skills in both research and analysis. I would not be where I am, and this dissertation would not be what it is, without the benefits of his advice and guidance.

Chapter 1: Neural tube defects

1.1 Morphogenesis of the neural tube

The neural tube (NT) is the precursor to the vertebrate central nervous system (reviewed in Greene and Copp, 2009). Most of the NT forms by primary neurulation, which for our purposes will be considered synonymous with 'neurulation' (Figure 1A). Briefly, neurulation begins with induction of the neural plate, a thickened region of dorsal ectoderm, above the notochord. The cells of the neural plate subsequently undergo convergent extension (Figure 1B), converging upon and intercalating along the plate midline, causing the neural plate to both lengthen and narrow. The lateral edges of the neural plate, termed the neural folds, elevate and converge above the midline through a combination of cell shape changes (cf. apical constriction at hinge points), cell rearrangements and division (Ezin et al., 2009), and extrinsic forces from the adjacent epithelium (Alvarez and Schoenwolf, 1992). The neural folds then meet and fuse to close the neural tube. Fusion is initiated at multiple discrete sites and proceeds outward to 'zipper' the neural tube closed, with the cranial region closing around day 26 and caudal around day 28 in human development.

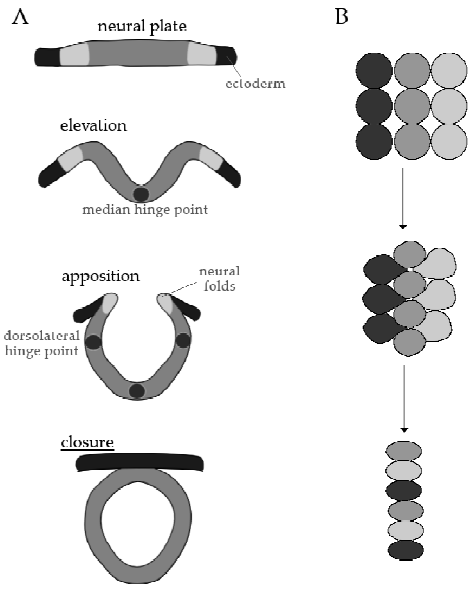

Figure 1. Overviews of A) primary neurulation and B) convergent extension.

Patterning of the neural tube is coordinated primarily by interactions between canonical Wnt, Shh, and BMP (TGF-β superfamily) signaling pathways. BMPs act as dorsal patterning signals, originating with the roofplate and defining dorsal cell identity (Chesnutt et al., 2004), while SHH signaling originates with the notochord and defines ventral cell identity (Roelink et al., 1995). BMPs are also required to maintain the roofplate after its formation, and pattern the expression of Wnt signaling components along the dorsal-ventral axis (Chesnutt et al., 2004). Members of the Wnt pathway also promote cell proliferation and dorsal identity, partly through regulation of BMP signaling (Bonner et al., 2008), and antagonize Shh signaling to constrain ventral cell patterning (Yu et al., 2008). Generally, deficits in dorsal identity (Chesnutt et al., 2004)

and/or aberrant expansion of ventral patterning (Ybot-Gonzalez et al., 2002) result in a failure of neural tube closure. Other processes intimately linked to neural tube specification and closure include primary cilium morphogenesis and intraflagellar transport, required for Shh signaling (Huangfu et al., 2003), and noncanonical Wnt signaling (planar cell polarity, or PCP), which orchestrates convergent extension (Wallingford and Harland, 2002). Wnt signaling, canonical and non-canonical alike, is described in further detail below (Section 1.3.2).

1.2 Neural tube defects

1.2.1 NTD phenotypes

Neural tube defects (NTDs) arise from defective closure of part or all of the neural tube, and are among the most common of human birth defects. Here we focus on failures of primary neurulation, which are classified into the phenotypes spina bifida (SB), the failure of closure in caudal regions of the neural tube; anencephaly (AN), failure of closure in the cranial neural tube; and craniorachischisis (CR), failure along the length of the neural tube.

Anencephaly is characterized by the absence of the forebrain and cranial vault; infants with AN may survive to term, and for hours to days after birth, but the phenotype is ultimately lethal. In lieu of the forebrain is the *area cerebrovasculosa*, a mass of fibrovascular tissue containing some neural elements (Elwood et al., 1992). The

midbrain and cerebellum may or may not develop, while the brain stem is generally present but incomplete. Evidence suggests that in some cases, the forebrain may initially develop but degrade over time (Cafici and Sepulveda, 2003), referred to as the acrania-anencephaly sequence. In model organisms, the most prevalent cranial NTD is exencephaly, in which cerebral tissue overgrows through the open cranial vault; the exposed tissue frequently degrades if gestation is continued, producing an anencephalic phenotype (Matsumoto et al., 2002).

Spina bifida encompasses lesions arising from defective closure of the caudal neural tube. These can be subdivided into a number of phenotypes based upon the tissue types present in the lesion, including meninges, neural tissue, and/or fatty tissue. Spina bifida can also be categorized by lesion level (e.g. thoracic, lumbar). This phenotype is frequently accompanied by tethering, the abnormal attachment of the spinal cord to adjacent tissues, and hydrocephalus, the obstruction of cerebrospinal fluid drainage from the brain (Greene and Copp, 2009). Unlike other NTD phenotypes, SB cases are generally viable. Spina bifida co-occurs infrequently with AN; only some 20% of mouse models exhibit both phenotypes (Harris and Juriloff, 2007).

1.2.2 Prevalence of NTDs and therapeutic interventions

Between 1996 and 2000, the overall global prevalence of AN was 2.1 per 10,000 births, and 3.4 per 10,000 births for SB (Lie, 2006). However, prevalences varied widely

between regions, ranging from 0.4 per 10,000 AN cases in Spain to 13.5 per 10,000 in Mexico, and 2.3 per 10,000 SB cases in the U.S. to 14.4 per 10,000 in Mexico. Some variability in prevalences can be explained by confounding factors in ascertainment, such as the inclusion or exclusion of terminations and the use of a population-based or hospital-based study design; however, genetic background and regional environmental factors also vary by region and/or ethnicity. For example, consistently high prevalences have been recorded for regions including China, India, Mexico, and the Middle-East (Lie, 2006). One of the highest incidences of NTDs recorded is ascribed to a four-county region in Shanxi Province, with 1.4 affecteds per 100 births (Li et al., 2006). All told, NTDs collectively represent a considerable public health burden around the globe.

The U.S. fully implemented mandatory fortification with folic acid in 1998. Subsequently, the prevalences of AN and SB have declined by upwards of 20% and 34% respectively (Canfield et al., 2005). Folic acid supplementation, whether by maternal vitamins or through fortification, is the only known preventative strategy for NTDs. Furthermore, while SB may be corrected with postnatal or even *in utero* surgery, prevention is the only treatment for more severe NTD phenotypes. Recent studies indicate that the U.S. population is now largely folate-replete, and no longer exhibit association between NTD risk and maternal folate intake (Mosley et al., 2009). This

suggests that folate supplementation has largely succeeded in preventing folate-sensitive NTDs, and new strategies will be needed to address the remainder of cases.

Using data from the National Birth Defects Prevention Study (1997-2007), Agopian et al. examined the contributions of risk factors in the fortified U.S. population for both AN and SB (Agopian et al., 2013). They found that overall, some 60% of population NTD risk remains unexplained. Major known risk factors included obesity (2-10%), Hispanic ethnicity (8-15%), low dietary folate intake (4-10%), and female infant sex (8%, AN only). Their data also shows that while family history of NTDs increases an individual's risk considerably (odds ratio 7-9), as a population risk factor, family history explains only some 2% of risk. However, under the generally accepted oligogenic model of NTD etiology, a portion of that unexplained risk may be ascribable to genetic factors which are not necessarily detrimental except in the right (or wrong) combinations.

1.3 Genetic processes contributing to NTD etiology

NTD etiology is known to have a strong genetic component. Although most human cases are sporadic, with only one affected in a family, siblings of a case are on average 32 times more likely to be affected than the general population (Elwood et al., 1992) . However, association studies in humans have had limited success in identifying and replicating genetic associations of candidate genes (reviewed in Au et al., 2010; Juriloff and Harris, 2012). Difficulties of studying NTDs in humans include the low

recurrence rates and lethality of most phenotypes; small sample sizes and poor statistical power for many studies; and heterogeneity of phenotypes and populations. Most established NTD candidate genes have been identified through animal models, particularly the generation of knockout strains in mice.

Using mouse models, over 240 genes have been implicated in NTD risk, a list which is not yet considered exhaustive (Harris and Juriloff, 2007, 2010). That said, many candidate genes fall into a central set of pathways and processes, including but not limited to Hedgehog signaling, primary cilium morphogenesis and function, chromatin structure, regulation of actin cytoskeleton dynamics, and noncanonical Wnt or planar cell polarity (PCP) signaling. As the most directly relevant to our research and findings, we focus here on folate metabolism, Wnt signaling, and chromatin modifications.

1.3.1 Folate metabolism

The single human SNP most robustly associated with NTDs is rs1801133, in 5,10-methylenetetrahydrofolate reductase (*MTHFR*). This SNP causes an Ala 222 > Val substitution which produces a thermolabile form of the protein, and has been estimated to explain up to 19% of NTDs, including a majority of folate-sensitive NTDs (Ou et al., 1996). Other variants in *MTHFR* and in other genes relating to folate metabolism, such as cystathionine β-synthase, thymidylate synthase, and methionine synthase reductase, have also evidenced modest contributions to NTD risk (Shaw et al., 2009a).

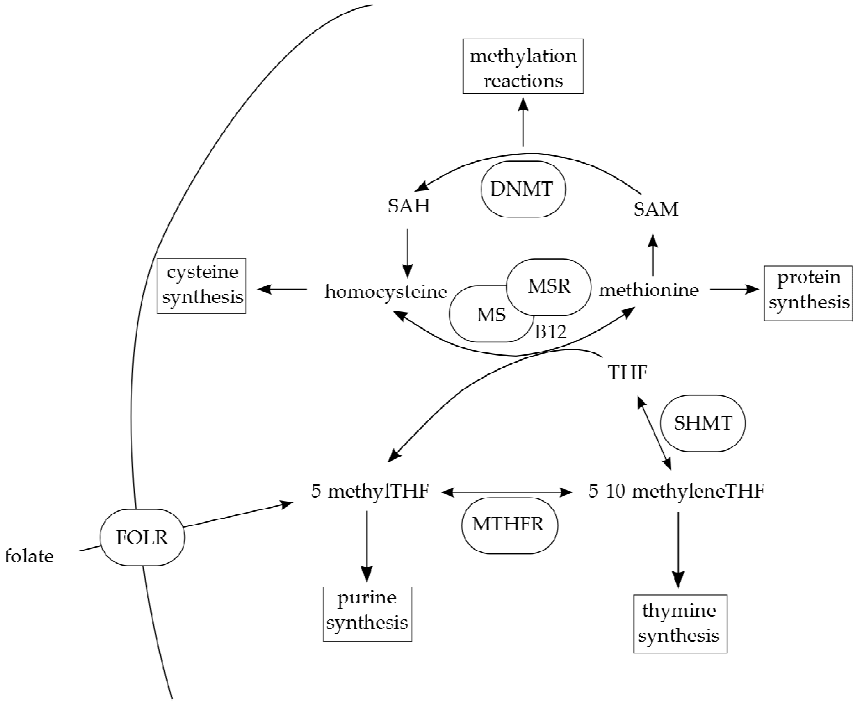

Figure 2. Overview of folate metabolism.

Folate derivatives are required for nucleotide synthesis (purines and thymine), protein synthesis (cysteine and methionine), formate metabolism, and methylation reactions; defects in all of these processes have been implicated in NTD etiology either directly or indirectly (reviewed in Beaudin and Stover, 2009). Deficiency in thymidylate synthesis, for example, is characteristic of the *splotch* mouse strain, a model system with completely penetrant NTDs in homozygotes (Fleming and Copp, 1998). In an example of more indirect connections, nucleotide and protein synthesis are essential for cell proliferation, and defects of proliferation can impair closure (Copp et al., 1988). Finally,

one-carbon metabolism and methylation reactions have become a topic of particular interest in NTD research, discussed in more detail below (Section 1.4.1).

1.3.2 Wnt signaling

As described above, Wnt signaling plays important roles in the morphogenesis of the neural tube. In the canonical branch, secreted Wnt ligands bind to Frizzled (Fzd) family receptors and Low density lipoprotein receptor-related protein (LRP) co-receptors (reviewed in Nusse, 2012). This leads to formation of a complex involving Disheveled (Dvl) family members, Axin, and Glycogen synthase kinase 3 (GSK3). The complex inhibits phosphorylation of β-catenin by GSK3, allowing β-catenin to translocate to the nucleus and convert transcription factors of the TCF/LEF family from repressors to activators of Wnt target genes. The different activities of canonical Wnt signaling in NT development can largely be segregated by ligand. Mitogenic effects are driven by ligands WNT1 and WNT3A (Megason and McMahon, 2002), while other family members are specific to dorsal (WNT3) or ventral (WNT5A, WNT7A/7B) regions, and some (e.g. WNT4) are expressed more broadly (Parr et al., 1993).

Non-canonical Wnt signaling determines planar cell polarity: the organization of cells within a plane of tissue, orthogonal to the apical-basal axis (reviewed in Gray et al., 2011). Core PCP genes are conserved from invertebrates through mammals, and include Frizzled (FZD3 and FZD6), cadherin, EGF LAG seven-pass G-type receptors (CELSR1-3),

Van Gogh-like (VANGL1-2), Dishevelled (DVL1-3), Prickle (PRICKLE1-4), and ankyrin repeat domain 6 (ANKRD6). Mouse knockouts of many of these genes have defects in convergent extension which result in NTDs. PCP signaling is also essential for ciliogenesis and Shh signaling (Park et al., 2006).

Briefly, in vertebrate PCP signaling (reviewed in Gray et al., 2011), Wnt ligands (WNT5/11) bind to FZD receptors which complex with CELSR proteins to activate DVLs. CELSR proteins may also complex with VANGL family members, which antagonize the activating complex, likely required for propagating polarity among adjacent cells. Activated DVL complexes with disheveled associated activator of morphogenesis 1 (DAAM1), Rac, and Rho, leading to activation of downstream effectors which regulate actin cytoskeleton dynamics and drive changes in cell shape. In model systems where defective PCP signaling impairs convergent extension, the neural tube remains too wide for the folds to meet and fuse (Wallingford and Harland, 2002). Several studies on humans have implicated rare variants or putative mutations of PCP genes in human NTD etiology, but have not investigated the contributions of more common alleles (reviewed in Juriloff and Harris, 2012).

1.4 Environmental factors contributing to NTD risk

A number of non-genetic factors have been implicated in NTD pathogenesis. Folate deficiency is among the best-known, but other dietary deficiencies which may

contribute to risk include methionine (Shoob et al., 2001), vitamin B12 (Ray et al., 2007), zinc (Velie et al., 1999), selenium (Cengiz et al., 2004), and inositol (Cockroft et al., 1992). In addition to deficiencies, a number of early-pregnancy exposures have been implicated in NTD risk, including alcohol (Liu et al., 2009; Wang and Bieberich, 2010), cotinine (Shaw et al., 2009b), a variety of metals (Huang et al., 2011; Jin et al., 2013), medications such as valproic acid or trichostatin A (Menegola et al., 2005), and the mycotoxin fumonisin (Gelineau-van Waes et al., 2005; Missmer et al., 2006). Aspects of maternal health and metabolism, such as obesity and diabetes (Hendricks et al., 2001; Moore et al., 2000) or first-trimester illness and fever (Shaw et al., 1998) also associate with greater risk of NTDs. As with genetic contributions to NTD pathogenesis, the array of implicated environmental factors is large and complex, with many opportunities for interactions; here we focus on two elements, one-carbon metabolism and exposure to the mycotoxin fumonisin.

1.4.1 One-carbon metabolism, methylation reactions, and chromatin structure

The folate derivative 5-methyltetrahydrofolate is necessary for the regeneration of methionine from homocysteine (Figure 2). Methionine, in the form of S-adenosylmethionine, is the carbon donor for most methylation reactions in the cell (Kozbial and Mushegian, 2005). After methyl transfer, S-adenosylhomocysteine is converted to homocysteine, allowing the cycle to continue. Thus, folate metabolism is

intimately connected to methylation reactions, and the aberrant methylation of DNA, histones, and other molecules may contribute to NTD pathogenesis.

Support for this hypothesis comes from a number of directions. In addition to folate deficiency, methionine (Shoob et al., 2001) and vitamin B12 (Ray et al., 2007) deficiencies have been associated with increased NTD risk, B12 being another essential cofactor for methionine synthase. The HDAC inhibitors valproic acid and trichostatin A (Menegola et al., 2005) are both NTD-inducing teratogens. Knockout mice lacking the DNA methyltransferases *Dnmt1* (Li et al., 1992) or *Dnmt3b* (Okano et al., 1999) exhibit NTDs, and family member *Dnmt3a* has been shown to regulate neural and neural crest cell fate (Hu et al., 2012). Additionally, hypomethylation of brain tissue (Chen et al., 2010) and generally aberrant DNA methylation profiles (Chang et al., 2011) have been observed in human NTD cases, with a suggestion of greater hypomethylation correlating to more severe NTD phenotypes. Related to this, epigenetic factors have been suggested as an explanation for the female excess observed in AN (Gelineau-van Waes and Finnell, 2001); methylation of the second X chromosome causes a greater demand for methyl groups during development in females than in males. However, it remains to be determined whether aberrant methylation or other epigenetic modifications actively contribute to NTD pathogenesis in humans, and if so, what proportion of risk they may explain.

1.4.2 Fumonisin exposure and sphingolipid metabolism

1.4.2.1 The mycotoxin fumonisin and NTD risk

Fumonisins are a class of mycotoxins produced by *Fusarium* species, particularly *F. verticillioides* (reviewed in Stockmann-Juvala and Savolainen, 2008). Although 15 different fumonisins have been identified, fumonisin B_1 (FB1) is the most prevalent by far, and for the purposes of this paper will be considered "fumonisin". A study of global occurrence detected FB1 in 50-60% of food and feed materials worldwide, with the greatest prevalence in South American samples (Schatzmayr and Streit, 2013). Animal studies have associated FB1 with equine encephalomalacia, hepatic cancer, pulmonary oedema, liver damage, and kidney damage (reviewed in Stockmann-Juvala and Savolainen, 2008). Finally, and of most relevance to this research, some mouse strains exhibit NTDs when treated with FB1 (Gelineau-van Waes et al., 2005). However, at present little is known about fumonisin toxicity in humans, or its contribution to human NTD risk.

Evidence connecting fumonisin to human NTD pathogenesis is largely circumstantial. Regions in China, South Africa, and Latin America with high maize consumption also exhibit high NTD prevalences (reviewed in Marasas et al., 2004). Additionally, and maize tortilla consumption has been significantly associated with NTD risk (Missmer et al., 2006). These data, although circumstantial, support a closer investigation of FB1 as a potential risk factor for human NTDs.

1.4.2.2 Inhibition of sphingolipid biosynthesis by FB1

Fumonisin is known to competitively inhibit ceramide synthases due to structural similarity with their substrates (Figure 3). These enzymes (CERS1-CERS6 in humans) attach fatty acid groups to sphingoid bases to produce ceramides, which are key intermediates in sphingolipid metabolism (reviewed in Pewzner-Jung et al., 2006). Inhibition of ceramide synthase results in an excess of simple sphingolipids sphingosine (So) and especially sphinganine (Sa), and concomitant depletion of ceramide (Cer) and its derivatives (sphingomyelin, glycosphingolipids).

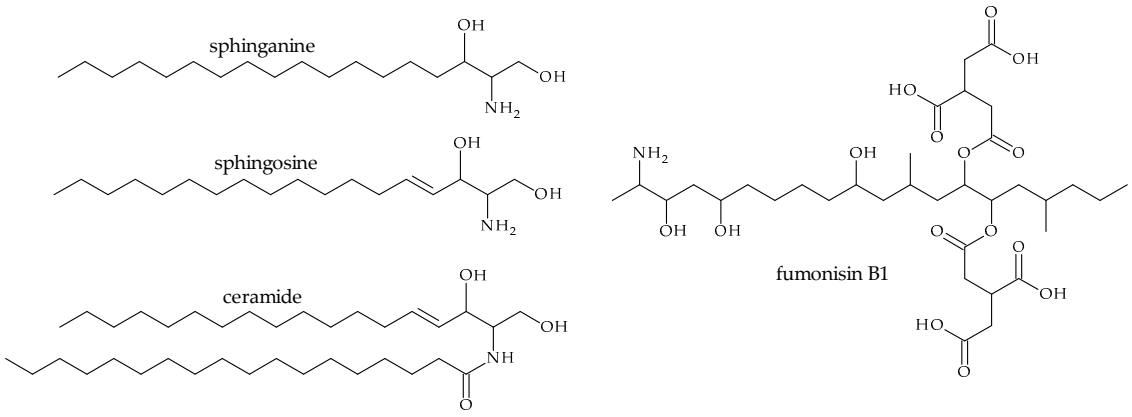

Figure 3. Structures of key sphingolipids and fumonisin B1.

Sphingolipids are important membrane components, but ceramide, Sa, and So are also bioactive lipids with key regulatory functions (reviewed in Mencarelli and Martinez–Martinez, 2013). Sphingosine promotes growth arrest, while its phosphorylated form (So1P) promotes proliferation. Conversely, Cer promotes apoptosis, and ceramide-1-phosphate (Cer1P) promotes cell survival; implicitly,

maintaining the balance between sphingolipid species is critical to cell function and survival. Additionally, ceramide has been implicated in primary cilium morphogenesis (Wang et al., 2009), which is necessary for Hedgehog signaling. Sphingolipids are also involved in the regulation of differentiation and cell migration, and can act as intracellular messengers, binding to specific G-protein-coupled receptors (S1PR1-S1PR5) on the surfaces of cells (reviewed in Maceyka et al., 2012). These receptors have been implicated in a number of processes and diseases, from immune responses to diabetes. Finally, So1P has been shown to inhibit HDAC1 and HDAC2, suggesting a role for simple sphingolipids in regulating epigenetic modifications.

1.5 A model of differential susceptibilities: LM/Bc and SWV mice

Human NTDs are not generally detected until much later in development, after the germ layers have differentiated and exposed tissues have had opportunity to degrade. This lag between defect formation and discovery has negative repercussions for retrospective studies, when exposures which contributed most to pathogenesis occurred weeks or months before ascertainment. It impairs epigenetic and gene expression experiments, as the most proximal tissues – the neural plate and surrounding elements – cannot be readily obtained; and particularly for open NTDs, such as AN, even differentiated neural tissues may be absent or degraded about the defect site. Finally, the genetic heterogeneity of human NTDs, described briefly above, can be a

confounding factor in its own right. Model organisms thus occupy a critical niche in NTD research, providing tractable systems in which these issues can be minimized.

In collaboration with the Gelineau-van Waes lab at Creighton University, we used the LM/Bc and SWV mouse strains as a model of the interactions between genetic and environmental risk factors. These strains are relatively closely related, with LM/Bc mice having been outcrossed to SWV mice in 1968; inbreeding was resumed with F1 individuals who were 1/8 SWV and 7/8 LM/Bc, and the strain has since been maintained as a closed, inbred colony (Harris et al., 1984). Yet despite this shared heritage, these strains have strikingly different responses to environmental factors. SWV mice are highly susceptible to NTDs induced by anticonvulsant drugs (Finnell et al., 1988) and elevated maternal body temperature (Finnell et al., 1986), while LM/Bc mice are highly susceptible to NTDs induced by the mycotoxin fumonisin (Gelineau-van Waes et al., 2005). Each strain is also relatively resistant to factors the other is sensitive to. These mirror-image responses constitute an excellent model system for studying genetic modifiers of NTD risk.

Chapter 2: The contributions of genetic factors to NTD etiology

Association studies in humans have had limited success in identifying and replicating genetic association of candidate genes and variants (reviewed in Au et al., 2010; Juriloff and Harris, 2012). This difficulty arises in part from the genetic heterogeneity of NTDs, exemplified by the over 240 candidate genes identified from mouse models (Harris and Juriloff, 2007, 2010). However, though the accepted model for NTD etiology is multifactorial, involving both genetic and environmental risk factors, interactions between risk factors remain relatively understudied in the field. Here we have identified candidate gene-gene interactions which may contribute to NTD pathogenesis using whole-exome sequencing in a multiplex human family and in a mouse model of differential NTD susceptibilities. Our data underscore the importance of genetic background to the etiology of highly penetrant as well as sporadic NTDs.

2.1 Oligogenic etiology in humans

We performed whole-exome sequencing on all members of a multiplex NTD family, including both parents, three anencephalic children, and one unaffected child. We hypothesized that the multiple occurrence of anencephaly would be due to a single highly penetrant genetic lesion homozygous in affected individuals and heterozygous in unaffected family members. However, we found no variant that fit this model. Instead,

we identified a collection of heterozygous rare and common variants within PCP signaling genes which were enriched in the affected offspring. This is to our knowledge the first study to take a cumulative, pathway-level approach for identifying potential interacting variants which is inclusive of both rare and common human alleles.

2.1.1 Methods

2.1.1.1 Sample collection and clinical information

The Hereditary Basis of Neural Tube Defects study at Duke University Medical Center has been enrolling participants since 1996 and the cohort consists of over 5,000 individuals with all types of NTDs (Section 1.2.1) and their family members. The study includes both singleton and multiplex families. Ascertainment of participants occurs across the United States from a variety of sources including referral by genetic counselors, physicians, support groups and self-referral in response to advertisements by the study team. We selected a single AN multiplex NTD family for whole-exome sequencing because of the availability of DNA from both parents, an unaffected sibling, and three separate AN pregnancies.

The family was of European ethnicity, with no prior history of NTDs, and affecteds included male and female individuals. Fetal tissue and/or cord blood samples were collected by the participant's physician at the time of termination or delivery for all four offspring. Blood samples were obtained from both parents. A detailed family

history was obtained, and medical records including ultrasound reports and images were collected to confirm the diagnosis; relevant clinical information is given in Table 1. This study was conducted under the oversight of Duke University Medical Center Institutional Review Board and informed consent was obtained from all participants.

Table 1. Clinical information for included pregnancies in the multiplex family.

	A1	A2	A3	U
Phenotype	AN	AN	AN	normal
Sample type	tissue	tissue	FFPE tissue	cord blood
Folate suppl.	no	no	yes	yes
Maternal multivitamin	no	no	yes	yes
Other supplementation	---	---	inositol	B complex, C, D, E, calcium, coenzyme Q10, alpha-lipoic acid
Maternal BMI	normal	normal	normal	normal
Other medication	Zoloft	Zoloft	Zoloft	Zoloft, antibiotics

2.1.1.2 Exome library preparation and sequencing

DNA was extracted from fetal tissue (affecteds A1 and A2), cord blood (unaffected), and peripheral blood (parents) by the Duke University DNA Bank and Tissue Repository on a Gentra Automated DNA Extractor using Puregene chemistry. Only formalin-fixed, paraffin embedded (FFPE) slides were available for affected A3; DNA was extracted from the FFPE slides by Genewiz according to their in-house protocol, and also by QIAamp DNA FFPE Tissue kit (Qiagen) for validation

experiments. Protein-coding regions were enriched for all samples using the TruSeq exome capture kit (Illumina, Inc., San Diego, CA), which targets 201,121 exons in 20,794 genes. Paired-end 100 nt reads were generated using an Illumina HiSeq instrument. Sample libraries for the father, affected A1, and affected A2 were generated at the Duke Center for Human Genetics Molecular Genomics Core facility and sequenced by the David H. Murdock Research Institute. Both library preparation and sequencing were performed by GENEWIZ, Inc. for all other samples.

2.1.1.3 Alignment and variant calling

Raw reads were aligned to the NCBI human reference genome build 37.1 using Burrows-Wheeler Aligner (Li and Durbin, 2009). Unmapped and duplicate reads were removed using samtools (Li et al., 2009) and PICARD (Wysoker et al., 2009) respectively. Local realignment around indels and base quality score recalibration were performed using Genome Analysis Toolkit (GATK) version 2.5 (DePristo et al., 2011). Single nucleotide variants were called simultaneously across all samples using the Unified Genotyper in GATK, constrained to 'target intervals' within 400 nt of any capture probe. The default sample contamination parameter (5%) was used for all samples except affected A2, which preliminary analysis determined was contaminated with maternal DNA, most likely at the time of termination. The contamination threshold for that sample was set at 10% after evaluation of maternal allele overrepresentation at

parentally homozygous sites; this corrected most erroneous genotype calls. Variants were phased for each child with the PhaseByTransmission tool in GATK. Finally, low-quality variants were filtered out according to the Broad Institute's recommendations for small sample sets as given in GATK documentation (Van der Auwera et al., 2013). Variants were annotated against human RefSeq genes using ANNOVAR (Wang et al., 2010) with a neargene threshold of 2 kb.

2.1.1.4 Variant segregation analysis

Sites where affecteds A1 and A2 both called homozygous for the minor allele, but not either parent or the unaffected, were considered to potentially segregate with phenotype. Due to expectation of a large detrimental effect, variants were further constrained to non-synonymous common SNVs and all SNVs with a minor allele frequency of less than 5%. Data was reviewed manually at these sites to exclude variants with low coverage or nearby indels affecting genotype calls.

2.1.1.5 Haplotype transmission analysis

Sites heterozygous in only one parent were used to construct haplotype blocks across each covered gene. Haplotypes for children A2, A3, and U were scored on their differences from to the corresponding haplotype of child A1, with 0 being an exact match. No penalty was assigned for missing data. We initially filtered on genes showing dual segregation of parental haplotypes with affection status, allowing up to

20% mismatch to accommodate possible calling errors. Subsequently, we filtered on genes where only one parental haplotype segregated with affection status, and followed up on variants transmitted by the opposite parent in those and other interacting genes.

2.1.1.6 Validation of novel variants

Primers for the validation of novel variants in *PAX6* and *TEAD1* were designed using Primer3Plus (Untergasser et al., 2007). Primer sequences are given in B.1 Human PAX6 and TEAD1 validation primers. Both products were amplified using HotStarTaq (Qiagen) with a touchdown from 64 C to 58 C over the first seven cycles and all further cycles annealing at 58 C. Amplified products from all family members were Sanger sequenced by Genewiz (South Plainfield, NJ) to confirm variant authenticity and transmission.

2.1.1.7 Predictions of variant functional effects

In silico functional evaluation was performed for 30 variants in eight core PCP genes (*CELSR1*, *DAAM1*, *DVL1*/*2*/*3*, *LRP6*, *VANGL1*/*2*) and two transcription factors involved in neural tube development (*PAX6*, *TEAD1*). Scores from dbNSFP 2.0 were used for functional analysis of the sole missense variant, including scoring by Polyphen-2, SIFT, MutationTaster, MutationAssessor, and Likelihood Ratio Test (Liu et al., 2013). For all SNPs, we evaluated potential impacts on gene splicing, modeling our approach after that of Houdayer et al. (Houdayer et al., 2008). Briefly, splice site prediction

utilities NNSplice (Reese et al., 1997), ASSP (Wang and Marín, 2006), and NetGene2 (Hebsgaard et al., 1996) were used to predict splice sites within 100 nt of each SNP for both reference and variant alleles. Real and predicted splice sites were additionally scored using MaxEntScan under the Maximum Entropy model (Yeo and Burge, 2004). Variants with a change in score of 10% by any predictive utility were considered candidates for affecting splice site function. Additionally, Human Splicing Finder (HSF), which aggregates predictions from a number of tools (Desmet et al., 2009), was used to identify potential impacts upon putative splice sites and regulatory motifs within 25 nt of each SNV. Finally, likelihood of effect was scored for all variants using Spliceman (Lim and Fairbrother, 2012) and for exonic variants only with MutPredSplice (Mort et al., 2014). Results from all *in silico* tools were combined into best-estimate predictions of each SNV's possible functional effect.

2.1.2 Results

2.1.2.1 Sequence coverage and alignment quality

All samples showed excellent alignment to reference (Table 2). On-target percentages varied due to difference in shearing parameters, as evidenced by insert sizes, but most samples had good coverage of target intervals nonetheless. Less input DNA was available for affected A3, leading to lower coverage for that sample. Conversely, the unaffected was oversequenced (150 million raw reads vs. less than 100

million for other samples), and its greater duplication rate can be ascribed in part to the capture of duplicate fragments. All samples were very similar in base quality per cycle, read quality, sequence content, and GC content as reported by FastQC (Andrews, 2010) (data not shown).

Table 2. Overview of human exome coverage metrics. "Target" intervals defined as probe locations +/- 200 nt.

Sample	Mapped	Duplicate	On Target	Mean Target Coverage	Mean Insert Size
M	99%	19%	82%	18x	182
F	99%	11%	63%	33x	272
A1	99%	7%	64%	26x	278
A2	99%	10%	64%	30x	267
A3	99%	22%	97%	12x	223
U	99%	43%	86%	35x	202

2.1.2.2 Segregation of homozygous variants with phenotype

We found a total of 16 rare (MAF<5%) or nonsynonymous variants at which homozygosity segregated with phenotype (Table 3Table 1). Many were associated with odorant or taste receptor genes. None were associated with genes likely to play roles in neural tube closure.

Table 3. Rare and nonsynonymous variants at which homozygosity segregated with affection status. AAF, alternate allele frequency; PP2, Polyphen-2.

SNP	Gene	Class	AAF	PP2 Score	PP2 Rank
rs17868323	*UGT1A7*	nonsyn	56%	0.00	benign
rs9324624	*ARHGEF37*	nonsyn	48%	1.00	probably damaging
rs182152607	*SSPO*	intronic	1%	---	---
rs35083184	*OR4A5*	nonsyn	12%	0.63	possibly damaging
rs11246609	*OR4C46*	nonsyn	58%	0.00	benign
rs2278108	*CLP1*	upstream	6%	---	---
rs1193851	*PCNXL3*	nonsyn	34%	0.98	probably damaging
rs10772420	*TAS2R19*	nonsyn	50%	0.00	benign
rs17277522	*OR6S1*	nonsyn	45%	0.99	probably damaging
rs11622969	*OR6S1*	nonsyn	54%	0.09	benign
rs11622794	*OR6S1*	nonsyn	55%	0.02	benign
rs9624	*TPPP2*	nonsyn	12%	1.00	probably damaging
rs9890664	*OTOP3*	nonsyn	67%	0.00	benign
rs4071641	*C17orf99*	nonsyn	56%	0.00	benign
rs3744793	*USP36*	nonsyn	43%	0.01	benign
rs524625	*ANKRD5*	nonsyn	26%	1.00	probably damaging

2.1.2.3 Segregation of parental haplotypes with phenotype

We identified no potential candidate genes in which both parental haplotypes segregated with phenotype. However, several PCP genes (*DAAM1, DVL2*, and *DVL3*) showed segregation of one parental haplotype with affection status. As PCP signaling is known to be critical to neural tube development and closure, we went on to examine other planar cell polarity genes for additional potentially interacting variants.

2.1.2.4 Multiple heterozygosity across PCP genes

We observed no set of variants of different parental origin within any one gene that segregated perfectly with phenotype. However, we did identify a substantial number of variants within key PCP genes in the affecteds (Table 4). Affecteds A1 and A2 both inherited a majority of variants, including rare variants from each parent. Affected A3 inherited only maternal rare variants, but received many common variants from both parents. Conversely, the unaffected inherited fewer PCP variants overall. Most strikingly, all affecteds had variants of different parental origins in at least two individual PCP genes, while the unaffected showed no such compounding.

Table 4. Multiple heterozygosity for SNVs in PCP genes. AAF, alternate allele frequency; MS, missense; SG, stopgain; S, synonymous; het, heterozygous genotype; ---, homozygous reference.

Gene	SNP	Region	AAF	M	F	A1	A2	A3	U
CELSR1	rs1807585	intron	84%	---	het	---	het	---	het
CELSR1	rs9616008	intron	9%	het	---	het	het	---	---
CELSR1	rs3747252	intron	13%	---	het	---	het	---	het
DAAM1	rs1253006	intron	18%	---	het	het	---	het	het
DAAM1	rs8022614	exon (S)	39%	het	---	het	het	het	---
DAAM1	rs61984495	intron	16%	---	het	---	het	---	---
DAAM1	rs17096091	intron	16%	---	het	---	het	---	---
DAAM1	rs941883	intron	64%	---	het	het	---	het	het
DAAM1	rs941884	exon (S)	64%	---	het	het	---	het	het
DAAM1	rs2146012	intron	82%	---	het	het	---	het	het
DVL1	rs149812757	intron	1%	---	het	het	het	---	---
DVL1	rs148991554	intron	2%	het	---	het	het	het	het
DVL1	rs3855955	intron	29%	het	---	het	het	het	het
DVL2	rs72839770	intron	35%	---	het	het	het	het	---
DVL2	rs2074216	exon (S)	36%	---	het	het	het	het	---

DVL3	rs147410100	intron	1%	het	---	het	het	het	---
LRP6	rs1012672	exon (S)	7%	het	---	het	het	het	het
LRP6	rs140962861	exon (MS/SG)	0.01%	---	het	het	het	---	---
LRP6	rs2075241	intron	18%	---	het	het	het	---	---
VANGL1	rs3700672727	exon (S)	0.01%	---	het	het	het	---	het
VANGL2	rs116152559	5′ UTR	1%	---	het	---	het	---	het
VANGL2	rs17371401	intron	25%	---	het	het	---	het	---
VANGL2	rs140848403	exon (S)	0.4%	het	---	het	---	het	---
VANGL2	rs12087593	intron	27%	---	het	het	---	het	---
VANGL2	rs17380127	exon (S)	26%	---	het	het	---	het	---
VANGL2	rs17380141	exon (S)	26%	---	het	het	---	het	---
		heterozygous sites		7	19	21	17	16	11
		pct heterozygous		27%	73%	81%	65%	62%	42%

Of variants in these genes, only one altered protein sequence: rs140962861 in *LRP6*, with a minor allele frequency of 0.01% according to the US National Institutes of Health Heart, Lung and Blood Institute-sponsored Exome Sequencing Project (ESP) (Fu et al., 2013). This variant results in either an Arg 1136 > Trp amino acid substitution or a stopgain depending on transcript isoform, and was rated deleterious by four prediction utilities (Polyphen-2, SIFT, MutationTaster, and MutationAssessor).

2.1.2.5 Novel heterozygous variants in PAX6 and TEAD1

We also identified two novel variants, both heterozygous, in transcription factors *PAX6* and *TEAD1*. These are both established NTD candidate genes: *Pax6* mutant mice can exhibit exencephaly (Cattanach et al., 1996), as do *Tead1/2* doubly-null mice (Sawada et al., 2008). Both variants were validated by Sanger sequencing (data not shown; primers in B.1 Human PAX6 and TEAD1 validation primers). The *PAX6* variant was

transmitted to all three affecteds but not the unaffected, while the *TEAD1* variant allele was transmitted to A3 and the unaffected. Though the *TEAD1* variant did not segregate with phenotype, it may be a contributing factor in A3.

2.1.2.6 *In silico* predictions of variant functional effects

Of the 30 variants examined, 20 (67%) variants examined coincide with potential TF motifs, while 26 (87%) coincide with potential enhancer elements and could have some effect on expression of their associated genes. Seven variants are predicted by at least one of NNsplice, ASSP, NetGene2, MaxEntScan, or Human Splicing Finder to strengthen cryptic splice sites, while two are predicted to weaken known splice sites (Table 5).

Table 5. Predicted score changes for cryptic splice sites in proximity to variants.

SNP	Gene	Dist	Type	NNS	ASSP	NG2	MaxEnt	HSF
rs2074216	*DVL2*	+6	Acceptor	+26%	---	---	-4%	---
rs2074216	*DVL2*	+18	Acceptor	+32%	---	---	-41%	---
rs1012672	*LRP6*	+0	Donor	---	---	---	---	+67%
rs8022614	*DAAM1*	+2	Acceptor	---	---	---	---	+71%
rs10500764	*TEAD1*	+10	Acceptor	---	+20%	---	+75%	---
rs17380127	*VANGL2*	+12	Acceptor	---	---	---	+79%	---

All variants were predicted to affect some splicing-regulatory motif in some fashion. In total, 6 of 9 exonic variants were predicted to affect an ESS or ESE element, with 17 of 20 intronic variants predicted to impact intronic silencing elements. Notably,

exonic variants in this set more often weaken enhancer elements and/or strengthen silencers, suggesting they may promote exon skipping, while more intronic variants weaken silencer motifs and could promote utilization of cryptic sites. Variants for which potential exon skipping, exon truncation, or intron inclusion could be extrapolated are given in Table 6.

Table 6. Summary of predicted variant effects on transcript structure.

SNP	Gene	Class	Alt AF	Predicted Impact on Sequence Features	Hypothesized Impact on Protein
rs3747252	*CELSR1*	intronic	13%	broken ISS, new predicted acceptor	exon 2 may begin early, stopgain
rs9616008	*CELSR1*	intronic	9%	new ISS	may skip exon 3, stopgain in exon 5
rs8022614	*DAAM1*	syn	39%	new ESS	may skip exon 6, stopgain in exon 7 may use alternative termination in exon 6
rs941883	*DAAM1*	intronic	64%	new ISS	may skip exon 13, stopgain in exon 14
rs941884	*DAAM1*	syn	64%	broken ESE	may skip exon 13, stopgain in exon 14
rs148991554	*DVL1*	intronic	2%	new ISS	may skip exon 15; truncates at exon 14
rs2074216	*DVL2*	syn	36%	2 new predicted acceptors	may include intron 3, stopgain may create aberrant exon 4, stopgain
rs147410100	*DVL3*	intronic	1%	new ISS, new predicted acceptor	exon 11 may begin early, stopgain may skip exon 11, no frame change
rs140962861	*LRP6*	nonsyn	0.01%	broken ESE	may skip exon 16, no frame change
rs2075241	*LRP6*	intronic	18%	weakened acceptor	may skip exon 16, no frame change
rs370067272	*VANGL1*	syn	0.01%	broken ESE	may skip exon 7, stopgain in exon 8

rs116152559	*VANGL2*	UTR5	1%	broken ESE	may skip exon 2; no start codon
rs17380141	*VANGL2*	syn	26%	broken ESE	may skip exon 8; truncates at exon 7 may use alternate termination in exon 7
chr11:31823331	*PAX6*	intronic	---	new ISS	may skip exon 7; no frame change
rs3026384	*PAX6*	intronic	29%	broken ISS, new predicted acceptor	exon 8 may begin early; stopgain may include intron 7, stopgain
chr11:12946401	*TEAD1*	intronic	---	new ISS	may skip exon 11; no frame change
rs10500764	*TEAD1*	intronic	32%	broken ISS, new predicted acceptor	exon 6 may start early, stopgain

2.1.3 Discussion

We found no evidence for a single primary causal variant underlying the anencephaly cases in this multiplex family. Instead, we observed an aggregation of 26 heterozygous variants within eight key planar cell polarity genes (*CELSR1, DAAM1*, *DVL1/2/3*, *LRP6*, *VANGL1/2*), with a core set spanning *DAAM1, DVL1, DVL2, DVL3* and *LRP6 (*Table 4). These findings echo the broader complex etiology of NTDs in human populations. Most notably, the three AN affected individuals all exhibited variants of different parental origins in at least two individual PCP genes. Conversely, the unaffected did not inherit variants from different parents within any PCP gene, and also inherited fewer PCP variants overall.

The functional interactions of variants within NTD-relevant genes remain relatively understudied. Evidence does exist to support that genetic defects in PCP signaling may compound to critical effect, especially among DVL family members. It has been shown that while only *Dvl2*-null mice exhibit NTDs in the single-knockout case, heterozygous deficiency of *Dvl2* or *Dvl3* combined with complete knockout of the other exacerbates penetrance and phenotype severity (Etheridge et al., 2008; Hamblet et al., 2002). *Dvl3* deficiency can similarly compound with deficiency of *Vangl2* (Etheridge et al., 2008). Finally, there is functional redundancy among DVL proteins, with each able to rescue deficiency of at least one other (Etheridge et al., 2008). The effects of variants

in any given DVL gene are thus mediated by the abundance or deficit of other family members; in a case such as this, where variants are present in multiple DVL genes, their cumulative effects may overturn the normal functional redundancy of these proteins.

These variants may further compound with the variants we observed in *DAAM1* and *LRP6*, both of which are established NTD candidate genes and interact with DVLs. A diagram showing the interactions of these genes is given in Figure 4. DVLs are the key mediators of PCP signaling, but DAAM1 is essential for DVL-Rho complex formation and downstream activation of RhoA (Habas et al., 2001). Additionally, although LRP6 is best known as a co-receptor for canonical Wnt signaling, it has recently been shown to be a necessary part of the DVL-Rho complex as well (Gray et al., 2013). Finally, *VANGL* family members have an antagonistic interaction with the FZD-DVL complex (Shafer et al., 2011) which is also essential for DVL-Rho activity; disruption of *VANGL2* has been shown to impair RhoA expression and signaling (Phillips et al., 2005). Variants across these genes are likely to have a cumulative effect on RhoA activation PCP signaling, even if their effect sizes are individually small.

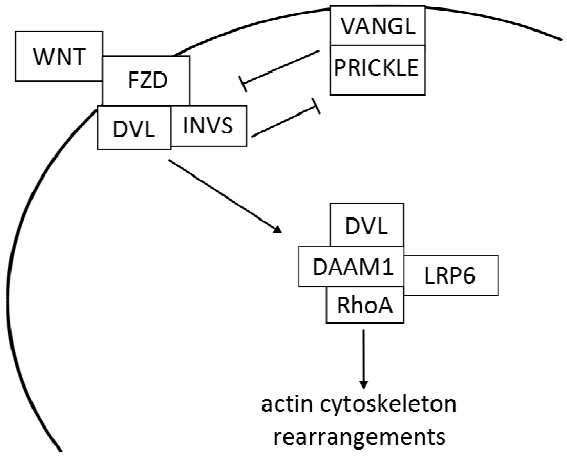

Figure 4. Interactions among selected members of the PCP signaling pathway.

In light of these interactions, it is perhaps telling that the unaffected offspring from our multiplex family had no compounding of different-origin variants within any single one of these genes. Furthermore, the unaffected inherited no variants in *DVL2* or *DVL3*, the two *DVL* family members most strongly implicated in NTDs. Conversely, the affecteds all exhibited mixtures of parental alleles both across *DVL* family members and within at least two key members of the DVL-Rho complex, which we posit collectively depressed PCP signaling below the level necessary for neural tube closure.

Additional variants in downstream effectors or essential transcription factors such as *PAX6* and *TEAD1*, both of which contained novel variants in this family, may have further compounded this deficit. PAX6 defines the future forebrain (Mastick et al., 1997) and is a marker for the dorsal neural tube regulated by Wnt signaling (Alvarez-

Medina et al., 2008). TEAD1 is a regulator of cell proliferation (Sawada et al., 2008) and a target of Hippo signaling, which crosstalks with both Wnt and Hedgehog signaling throughout development (Zhao et al., 2010). These variants are good candidates for additional interactors that may have contributed to the development of AN in this family.

Most of the variants we identified were synonymous or intronic, for which functional repercussions are difficult at best to predict. Nearly all of our variants were proximate to a splice junction, suggesting they may affect transcript splicing. Notably, studies of disease-causing mutations in BRCA1 (Liu et al., 2001) and ATM (Teraoka et al., 1999) suggest that as many as 50% of deleterious variants impact transcript structure, whether through directly affecting a splice site or by altering regulatory motifs. We did not have RNA available for this family and could not directly interrogate transcript structures. Instead, we used existing *in silico* utilities to predict effects on splice site recognition. We found these utilities to perform especially poorly with regard to donor sites. Although all utilities predicted roughly equal numbers of acceptors and donors within sequences, only half of real donor sites were identified in these genes, and those only by one utility (ASSP). Conversely, 85% of real acceptors were identified, most by all three utilities. This has poor implications for the ability of these tools to predict the impact of sequence variation on donor site recognition. We also were not able to predict

early termination of transcripts, though several variants are near alternative termination sites.

With the *in silico* utilities we used, two of our variants were predicted to negatively impact existing splice sites, while five were predicted to enhance potential cryptic splice sites (Table 6). In total, 6 of 9 exonic variants may promote exonic skipping through their effects on splicing regulatory motifs, including one which also weakened an existing acceptor, while 13 of 20 intronic variants may break intronic silencing elements, including 3 which also create potential cryptic splice sites (Table 6). Furthermore, most variants coincided with any of a number of TF binding motifs, and nearly all variants were located in potential enhancer elements, suggesting a potential for effects on expression even for variants that may not impact splicing. Ultimately, the functional effects of these variants can only be conclusively determined by examining transcript structure and expression levels of the associated genes. This is a critical aspect to be studied in the future, along with follow-up studies in larger cohorts, particularly focusing on total variant burdens across genes and pathways and addressing potential gene-gene interactions in NTD etiology.

Because the DNA for all the affecteds was derived from tissue from pregnancy terminations, an additional challenge in this analysis was maternal DNA contamination of affected A2. We found that increasing the contamination parameter in

UnifiedGenotyper successfully corrected the majority of variants. Miscalled sites that remained were sporadic, with few to none in any given gene; these we addressed by comparing haplotypes instead of individual variants, permitting a proportion of mismatches, and manually curating the resultant hits. This approach is robust against false negatives, though the standard caveat remains that a critical variant may have been located outside the regions exome sequencing covers. However, in this multiplex pedigree, we would expect any such variant to have a large effect and most likely occur in the coding region of a candidate gene, precisely what exome sequencing is suited to detect. These data underscore the importance of characterizing genetic background and genetic interactions even in multiplex families.

2.2 Genetic background as a modifier of NTD risk: evidence from a mouse model

In mouse models of NTDs, genetic background has been shown to influence NTD phenotype and penetrance. Examples include knockout of the gene *Men1*, which produces NTDs in 129S6/SvEv but not in C57BL/6 (Lemos et al., 2009), and differential phenotypes for crosses of the *splotch* mutation onto other strain backgrounds (Fleming and Copp, 2000). Genetic variants may additionally mediate environmental risk factors; most strikingly, mouse studies indicate that folate deficiency only causes NTDs when combined with a predisposing genetic background (Burren et al., 2010; Heid et al., 1992). Conversely, studies have also demonstrated that folate supplementation increases NTD

risk in some mouse models (Marean et al., 2011), an effect which may be particularly important to investigate in this era of mandatory fortification.

The role of genetic background in modifying NTD risk is strongly illustrated by the LM/Bc and SWV mouse strains, described in the previous chapter (Section 1.5). For this study, we focused on modifiers affecting susceptibility to FB1-induced NTDs in the LM/Bc strain. Fumonisins are competitive inhibitors of ceramide synthases (Section 1.4.2.2), which function in *de novo* sphingolipid biosynthesis (Figure 5), producing ceramide from simple sphingolipids such as sphingosine (So) and sphinganine (Sa). In the presence of fumonisin, levels of ceramide are depleted and simple sphingolipids become more abundant, particularly Sa.

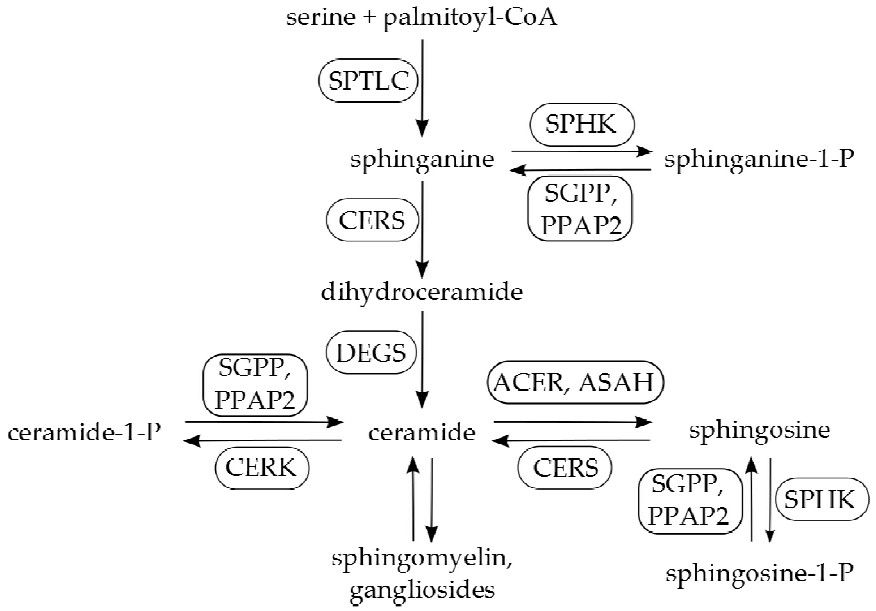

Figure 5. Overview of ceramide metabolism.

Ceramide derivatives are important components of lipid rafts that anchor membrane-bound receptors, including the folate receptor FOLR1. Indeed, experimental evidence shows that FB1-induced NTDs in LM/Bc embryos can be partially prevented with supplemental folic acid, and almost completely rescued with supplemental ganglioside GM1 (Gelineau-van Waes et al., 2005). However, the genetic factor or factors that underlie the strain's sensitivity to FB1 remain unknown. Here we use exome sequencing to identify variants in LM/Bc that are distinct from the SWV strain and may promote susceptibility to fumonisin-induced NTDs.

2.2.1 Methods

2.2.1.1 Exome library preparation, sequencing, and alignment

DNA from tail snips of representative LM/Bc and SWV mice, extracted using the Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI), was provided by the Gelineau-van Waes lab. Exonic regions were enriched for one mouse of each strain using the SureSelect XT Mouse All Exon kit (Agilent Technologies, Santa Clara, CA), according to manufacturer's recommendations. This kit targets 221,784 exons within 24,306 genes, with a total capture size of 49.6 MB. Prepared libraries were sequenced 2x100nt on an Illumina HiSeq at the David H. Murdock Research Institute. The resulting data was aligned against the NCBI37 mouse reference genome, the alignments cleaned, and variants called as previously described, except without use of

the contamination threshold parameter or allele phasing. Variants were annotated against mouse RefSeq genes as previously described. To reduce artifacts and false positives, variants were restricted to a read depth of at least 4, fewer than 10% of reads with mapping quality of 0, and a quality-by-depth of at least 7. Only sites homozygous for different alleles in LM/Bc and SWV were included in analyses.

2.2.1.2 Sanger validation of variants in *Cers1-Gdf1*

Primers were designed using Primer3Plus (Untergasser et al., 2007) to encompass the entire exon 8 of *Cers1-Gdf1*, with a nested sequencing primer; sequences are given in Appendix B: Primer Sequences. Primer oligos were ordered from Integrated DNA Technologies (Coralville, IA). Target regions were amplified using HotStarTaq DNA polymerase (QIAGEN, Germantown, MD) and sequenced on a 3730 DNA Analyzer using BigDye chemistry (Life Technologies, Carlsbad, CA). Variants were sequenced in both directions and in two mice of each strain.

2.2.1.3 Analysis of exome variants from AKR/J genome data

From the Mouse Phenome Database (Grubb et al., 2014), we determined the AKR/J strain shared the LM/Bc genotypes of sites in *Gdf1*. The Gelineau-van Waes lab at Creighton University also determined experimentally that AKR/J mice have moderate susceptibility to FB1-induced NTDs (see Section 2.2.3). To further evaluate genetic concordance between these strains, AKR/J whole-genome sequence data aligned to

GRCm38 was downloaded from the Wellcome Trust Sanger Institute Mouse Genomes Project (Yalcin et al., 2012). Exome target intervals were converted to GRCm38 coordinates using the liftOver webtool, and variants called over these regions as above. For analysis, variants were constrained on genotype as above, also requiring homozygous genotypes in AKR/J.

2.2.2 Results

2.2.2.1 Mouse exome sequence alignment

Both mouse exomes showed excellent alignment to reference (Table 7), low duplication rates, and high coverage of target regions. Both were also very similar in base quality per cycle, read quality, sequence content, and GC content as reported by FastQC (Andrews, 2010) (data not shown).

Table 7. Alignment metrics for mouse exome sequence data. Target interval defined as bait +/- 100 nt.

	Mapped	Duplicated	On Target	Mean Target Coverage
LM/Bc	99%	8%	71%	108x
SWV	99%	10%	70%	113x

2.2.2.2 Overview of genetic differences between LM/Bc and SWV mouse strains

A total of 362,528 variants were called from the exome data, of which 264,616 (73%) passed quality filters. Half of filtered variants (132,510) were homozygous and discordant between the two strains; LM/Bc was non-reference at 44% of those sites

(58,357), encompassing 4,256 genes. In total, 129 Gene Ontology (GO) biological process (BP) terms were found to be significantly enriched for these genes using g:Profiler (Reimand et al., 2011) with g:SCS multiple testing correction. The best-per-parent summarization of these enrichments is given in Table 8.

Table 8. Per-parent GO BP term enrichments for genes with variants in the LM/Bc strain. *P* values were corrected for multiple testing using g:SCS.

p-value	Term Name	Term ID	Genes
2.24e-40	single-organism process	GO:0044699	2555
3.46e-37	cellular process	GO:0009987	2728
1.84e-17	regulation of cellular process	GO:0050794	1821
2.41e-17	metabolic process	GO:0008152	1910
4.01e-16	localization	GO:0051179	974
4.99e-11	multicellular organismal development	GO:0007275	855
2.42e-10	cellular component organization or biogenesis	GO:0071840	897
1.14e-08	signaling	GO:0023052	1146
7.98e-08	signal transduction	GO:0007165	1063
3.29e-03	smoothened signaling pathway	GO:0007224	38
2.39e-02	regulation of protein kinase activity	GO:0045859	127
3.44e-02	regulation of insulin secretion	GO:0050796	41

Most enriched terms are broad metabolic, developmental, or cell communication processes, as the per-parent summary indicates. However, two specific enriched terms are notable: smoothened signaling pathway (*p*=3.29e-03), synonym of Hedgehog signaling, and nervous system development (*p*=6.01e-05). Although neural tube closure was not a significantly enriched term, many genes involved in neural tube morphogenesis and closure contribute to other aspects of nervous system development.

These findings suggest that LM/Bc is particularly distinct from SWV and from reference in genes relating to nervous system development.

2.2.2.3 Nonsynonymous variants in *Cers1-Gdf1*

Individually, the most notable sites were two missense variants in the *Cers1-Gdf1* bicistronic transcript (Table 9). Both are located in exon 8, in the pro-domain region of GDF1. These variants were validated by Sanger sequencing (data not shown). Functional impacts of amino acid substitutions were scored with Polyphen-2. As an additional indicator of effect, prevalence of the LM/Bc alleles across mouse strains was determined using the Mouse Phenome Database standard SNP variation query.

Table 9. Nucleotide and amino acid changes for variants in *Cers1-Gdf1*.

	Reference	LM/Bc	SWV	Prevalence of LM/Bc allele	Polyphen-2 score
rs32650885	A serine	A serine	T cysteine	9/17 strains (52%)	benign (0.00)
rs47220823	G alanine	A threonine	G alanine	1/23 strains (4%)	possibly damaging (0.66)

2.2.2.4 Summary of variants associated with sphingolipid metabolism genes

Of variant sites in LM/Bc not shared by SWV, a total of 271 were associated with sphingolipid metabolism genes, summarized by gene in Table 10. Particularly notable genes include *Degs2* (14 variants), *Acer2* (3 variants), *Acer3* (50 variants), *Cerk* (2 variants), and *Ppap2a* (3 variants); refer to Figure 5 for their roles in sphingolipid

biosynthesis. Although none of these variants were nonsynonymous, some may have regulatory and/or splicing effects that contribute to the strain's sensitivity to fumonisin.

Table 10. Summary by gene of LM/Bc-specific variants in sphingolipid metabolism genes. In parentheses are counts of nonsynonymous and total variants.

Acer2 (0/3)	*Acer3* (0/50)	*Aldh5a1* (1/51)
Arsi (1/11)	*Cerk* (0/2)	*Cers5* (1/11)
Cers6 (0/5)	*Degs2* (0/14)	*Gba2* (1/2)
Hexa (0/11)	*Kdsr* (0/2)	*Naga* (0/11)
Neu3 (3/23)	*Ppap2a* (0/3)	*Ppm1l* (0/13)
Ppt1 (0/7)	*Prkaa1* (0/1)	*Smpdl3b* (1/3)
St8sia2 (0/33)	*St8sia3* (0/1)	*Sumf1* (0/14)

2.2.2.5 Summary of variants in genes associated with neural tube phenotypes

Of variants in LM/Bc but not SWV, 2,467 sites were distributed across 134 genes associated with either abnormal neural tube morphology or exencephaly. The ten phenotype genes with the greatest number of discordant variants are given in Table 11. Notably, these include Wnt signaling co-receptors *Lrp2* and *Lrp6*. Phenotype genes not making the top ten overall, but containing high numbers of nonsynonymous variants are *Brca1* (9 NS, 22 total), *Fat4* (6 NS, 32 total), and *Hoxb2* (4 NS, 11 total). These genes and variants may be good candidates for modifiers affecting LM/Bc sensitivity to fumonisin.

Table 11. Summary of the top ten genes with variants in LM/Bc not shared by SWV. *All exonic variants in *C2cd3* annotated as amino acid substitution 'unknown'.

Gene	Variants	NS	Syn	Intronic	5'UTR	3'UTR
Pdgfrb	154	3	18	133	0	0
*C2cd3**	148	---	---	115	0	0
Tenm4	148	3	39	104	1	1
Gtf2i	128	0	9	119	0	0
Lrp2	127	5	10	112	0	0
Tcof1	118	18	9	89	1	0
Lrp6	114	5	10	98	0	1
Bbs7	112	4	14	85	1	7
Itga3	110	3	17	87	0	3
Tiam1	89	2	9	78	0	0

2.2.2.6 Concordance of LM/Bc variants with AKR/J genotypes

Of variants in LM/Bc but not SWV, 57,850 sites were also called homozygous for any allele in AKR/J. AKR/J shared the LM/Bc allele at 22,814 (39%) of these sites, and possessed a different allele at 35,036 (61%) sites (Table 12).

Table 12. Summary of LM/Bc concordance with AKR/J at variants not in SWV.

	All	Neural Tube	Exencephaly	Sphingolipid	Combined
Concordant Genes	1,892	50	25	8	67
Concordant Vars.	22,814	830	826	55	1,382
Discordant Genes	2,854	72	42	14	99
Discordant Vars.	35,036	1,129	445	216	1,380

Across genes of interest, the number of concordant variants essentially equaled the number of variants not also found in AKR/J. However, discordant variants did not follow the same distribution across these gene sets; particularly, there were more

discordant variants in sphingolipid metabolism genes, including *Acer2* (3 variants), *Acer3* (50 variants) and *Cerk* (2 variants). Discordant sites were also more numerous across neural tube morphology genes, including *Celsr1* (48 variants), *Fzd1* (4 variants), *Gli2* (1 variant), *Smo* (1 variant), *Folr1* (16 variants), *Dnmt1* (59 variants), and *Tead1* (11 variants). Interestingly, variants in *TEAD1* also showed significant association with sphingolipid metabolite levels in our related human study (Section 4.2.2.2). *Tead1* variants and others not shared by SWV or AKR/J mice are among the most likely to promote the greater susceptibility of LM/Bc embryos to FB1-induced NTDs.

2.2.3 Discussion

Exome sequencing of fumonisin-susceptible and -resistant mouse strains identified 132,510 homozygous sites with different alleles in the LM/Bc and SWV mouse strains. At 44% of those sites, LM/Bc had a non-reference allele; these 58,357 variants are candidate modifiers that may promote sensitivity to FB1-induced NTDs. Conversely, some among the 74,269 sites where SWV had a non-reference allele are likely to inform that strain's susceptibility to NTD-inducing teratogens other than fumonisin; however, analysis of those variants is beyond the scope of this thesis.

As fumonisins are competitive inhibitors of ceramide synthase, the two discordant, nonsynonymous variants in *Cers1-Gdf1* are of particular interest (Table 9). This gene is a bicistronic transcript (Figure 6A) which produces ceramide synthase 1

from the 5′ end (exons 1-7) and growth and differentiation factor 1, a TGF-β family ligand, from the 3′ end (exons 7-8). As other TGF-b family members, GDF1 contains a signal peptide, a pro-domain, and the active domain (Figure 6B) which is cleaved into the mature protein (Kingsley, 1994). After cleavage, the pro-domain peptide may form a complex with the mature protein, continuing to inhibit signaling.

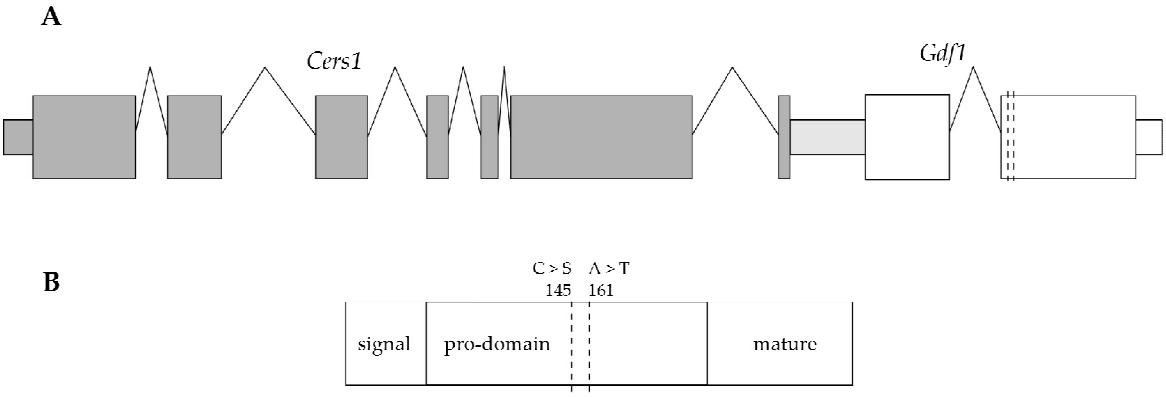

Figure 6. Location of strain-specific missense variants within A) *Cers1-Gdf1* bicistronic transcript and B) GDF1 protein structure.

Both of our missense variants are located in the pro-domain, and we hypothesize that the resultant amino acid differences may affect complex stability. Specifically, the cysteine residue in SWV may form a disulfide bond which, along with the hydrophobic alanine at the second site, promotes association. Conversely, the LM/Bc alleles code for polar acids serine and threonine that may promote complex disassociation and increase GDF1 activity. *Gdf1* knockout has been shown to increase *Cers1* expression (Bengtsson et al., 2008); accordingly, increased GDF1 activity can be expected to decrease CERS1

abundance, increasing susceptibility to the effects of fumonisin. *In situ* hybridizations performed by the Gelineau-van Waes lab show that exencephalic FB1-treated LM/Bc embryos have increased expression of *Gdf1* at E9.5 relative to controls (data not yet published), supporting this hypothesis.

Through the Mouse Phenome Database, we identified a second mouse strain, AKR/J, which shared alleles with LM/Bc at both *Gdf1* variant sites. The Gelineau-van Waes lab has performed fumonisin exposure experiments in this strain, and observed a 20% incidence of NTDs in AKR/J under the same conditions which produce 79% NTDs in LM/Bc (data not yet published). Exencephalic FB1-treated AKR embryos also show increased *Gdf1* expression by *in situ* hybridization, though not as strongly as their LM/Bc counterparts (data not yet published). These data indicate that these *Gdf1* variants contribute to but do not fully explain the susceptibility of LM/Bc mice to FB1-induced NTDs.

To identify other potential genetic modifiers of susceptibility to FB1, we contrasted our LM/Bc exome data with the Mouse Genome Project AKR/J genome. Interestingly, when considering only sites with different alleles in LM/Bc and SWV, we observed more discordance than concordance between LM/Bc and AKR/J in genes associated with sphingolipid metabolism and abnormal neural tube morphology (Table 12). Of particular note is *Folr1*, which contained 16 variants in LM/Bc not shared by

AKR/J. Some of these may act to reduce FOLR1 abundance; that FB1-induced NTDs can be rescued by folate suggests depletion of FOLR1 would increase susceptibility. This also echoes existing literature on the interactions between genetic background and folate deficiency as an NTD risk factor (Burren et al., 2010). Other variants in components of sphingolipid metabolism (*Acer2, Acer3, Cerk*), Wnt signaling (*Celsr1, Fzd1*), and Hedgehog signaling (*Gli2, Smo*), are also likely candidates for increasing susceptibility in the LM/Bc strain beyond that of AKR/J and SWV.

2.3 Conclusions: the multiple roles of genetic risk factors in NTD etiology

Genetic factors are an unquestionably strong component of NTD etiology. Loss of function of one gene may be sufficient to cause NTDs, as is seen in most mouse models. Alternately, NTD phenotypes may arise from interactions of lesions across multiple genes, as described by the oligogenic model generally accepted for human NTDs; our data indicate this to be the mechanism at work in the multiplex family described above (Section 2.1). Most importantly, we document that a combination of rare and common alleles, as opposed to rare variants or putative mutations with large effect sizes, may underlie NTD pathogenesis even in highly penetrant NTDs. The contributions of common variants and cumulative effect to human NTD pathogenesis remain generally understudied, and our findings underscore the importance of characterizing these in human NTD research.

Genetic risk factors also may not be sufficient to cause NTDs in and of themselves, but create a permissive background in which an environmental risk factor supplies the "second hit". This is most readily demonstrated with animal models, such as the differential sensitivities to FB1 we observed between LM/Bc, AKR/J, and SWV mice (Section 2.2). The contributions of fumonisins to human NTD pathogenesis remain to be determined but, combined with the human SNP association study in Section 4.2, our findings indicate that genetic background modulates the effects of FB1 in both mice and humans. These data demonstrate a need for further elucidation of gene-environment interactions in human NTD etiology, not only in the context of FB1, but for other environmental risk factors as well.

Chapter 3: Altered DNA methylation as a potential contributor to the etiology of human NTDs

At the interface between genes and environment, epigenetic marks, particularly DNA methylation, have recently become a key topic in NTD research. Folic acid, the only known preventative agent for NTDs, is an essential cofactor for methylation reactions (Beaudin and Stover, 2009). Deficiencies of other methylation cofactors, namely methionine (Shoob et al., 2001) and vitamin B12 (Ray et al., 2007), have also been implicated in NTD risk, while histone deacetylase inhibitors such as valproic acid can cause NTDs (Eikel et al., 2006). Simple sphingolipids, which are increased by fumonisin exposure, have also been identified to act as histone deacetylases (Hait et al., 2009).

The expression patterns of DNA methyltransferases 3a and 3b (*Dnmt3a, Dnmt3b*) in mouse embryos around the time of neural tube closure also have suggestive implications. *Dnmt3b* is strongly expressed in the very early embryo, but is replaced by *Dnmt3a* around E8.5-E10.5 (Okano et al., 1999; Watanabe et al., 2002), just before and during the time of neural tube closure. Around that time, *Dnmt3b* is expressed in neural ectoderm and the future forebrain, with *Dnmt3a* being ubiquitous; both are present in the embryonic ectoderm (Okano et al., 1999). Although *Dnmt3a* knockout mice gestate normally, *Dnmt3b* knockouts show normal development only prior to E9.5, with subsequent growth impairment and cranial neural tube defects (Okano et al., 1999). This

suggests that errors in DNA methylation, particularly at very early developmental stages, may indeed contribute to failure of neural tube closure.

Taken together, these relationships suggest a link between perturbed DNA methylation and NTD etiology. Slight but significant global hypomethylation has been observed in the in brain tissue of NTD cases, and the degree of hypomethylation may correlate with NTD phenotype severity (Chen et al., 2010). A second study documented significantly perturbed distributions of genomic methylation between cases and controls in multiple tissues (Chang et al., 2011). However, no functional evidence connecting altered DNA methylation with NTD etiology has been documented as of yet. It remains to be determined whether changes in methylation contribute directly to NTD development or are a side-effect of aberrant developmental processes.

3.1 Differential methylation in a monozygotic twin pair discordant for anencephaly

To reduce confounding variables, we selected a monozygotic (MZ) twin pair discordant for AN for this study. These twins were expected to be genetically identical, based on their MZ status, and had no observed *in utero* complications which might have contributed to the defect phenotype; thus, they provided an excellent context in which to identify possible relationships between DNA methylation and anencephaly. We performed exome sequencing and whole-methylome profiling on both twins, and included their parents and an unaffected sibling in validation experiments.

3.1.1 Methods

3.1.1.1 Sample collection and clinical information

This twin pair was selected for study because of their monozygotic twin status, implying identical genetics, combined with phenotypic discordance for anencephaly. They were collected as participants in the Hereditary Basis of Neural Tube Defects study as described in Chapter 2. Relevant clinical information is given in Table 13. Cord blood was collected immediately following delivery of the twins, both of which were live-born, and DNA was extracted by the Duke University DNA Bank and Tissue Repository using PureGene chemistry on a Qiagen Autopure instrument (Qiagen, Germantown, MD).

Table 13. Clinical details for the MZ twin pregnancy.

Twin pair gender	Female
Parental ethnicity	Caucasian
Maternal BMI category	Normal
Prenatal supplement use	Yes, pre- and post-conception
Relevant maternal exposures	Consumed wine and liquor pre-conception, smoked occasionally (<1 cigarette per day), took Claritin, used a hot tub pre- and post-conception, had mononucleosis just prior to conception.

3.1.1.2 Exome sequencing and analysis

Extracted DNA was sheared to 200 nt on a Covaris S2 Adaptive Focused Acoustic Disruptor and enriched for coding regions using the SureSelect Human All

Exon kit (Agilent Technologies, Santa Clara, CA) according to manufacturer's directions. Exome fragments were sequenced on an Illumina HiSeq as 100 nt paired-end reads. Reads were aligned to the NCBI human genome build 37.1, alignments refined, and variants called as previously described (Sections 2.1.1.3, 2.2.1.1). Only variants with a quality score of at least 50, a depth of at least five reads in each sample, and a minimum quality-by-depth score of 10 were included in analysis. Variant sites were mapped to their corresponding RefSeq and Ensembl genes using ANNOVAR (Wang et al., 2010) with a neargene threshold of 2 kb.

3.1.1.3 Methylation profiling by MBD-seq

Extracted DNA was sheared as above and enriched for methylated regions using the MethylMiner kit (Life Technologies, Grand Island, NY). Captured fragments were sequenced on an Illumina HiSeq as 100 nt single-end reads. Image analysis, base calling, alignment to the human genome build 37.2, and alignment cleaning were performed as above. Picard was used to determine read depths over all CpGs queried by the HumanMethylation450 microarray. Model-based Analysis for ChIP-Seq (MACS) was used to identify DMRs significant at p=1e-05 (Zhang et al., 2008). DMRs were annotated as for SNVs.

3.1.1.4 Methylation profiling by HumanMethylation450 array

Genomic DNA was bisulfite converted using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA) and run on HumanMethylation450 (450k) microarrays per the manufacturer's protocol (Life Technologies). Probes were excluded if missing data in any sample, if more than 80% of samples had a detection p-value greater than 0.05 for that probe, or if the probe had been previously identified as cross-reactive with other genomic locations (Chen et al., 2013). Peak correction was performed using *IMA* (Wang et al., 2012) to normalize the two different types of probes present on the chip. Color channel adjustment and quantile normalization were performed using *methylumi* (Bilke et al., 2012). Sites with a delta beta of 10% or greater between the twins were considered differentially methylated; these sites were annotated against Ensembl 72 genes, as above. Term enrichment analysis was performed using the g:Profiler webtool, with the developer's recommended g:SCS multiple-testing correction (Reimand et al., 2011). Sites were additionally annotated against ChIP-seq data for AP-2α in HeLa-S3 cells (ENCODE Project Consortium, 2011) and CDX2 in proliferating and differentiated Caco-2 cells (Verzi et al., 2010). Both datasets were downloaded from ChIPBase (Yang et al., 2013).

A second MZ twin pair, with twin-to-twin transfusion syndrome (TTTS), was also included in this experiment for comparison purposes. These samples were collected, extracted, and bisulfite converted as described above.

3.1.1.5 Validation of DM sites by Epityper assay and high-throughput bisulfite sequencing

Five DMRs from the MBD-seq data (in *CAMTA1, LRP1B, DAAM2, APC, GLI1*) were selected for validation in the original MZ twin pair by Sequenom EpiTyper assay on the basis of potential gene involvement in neural development and regulatory features in ENCODE data. The two most striking DMRs from the 450k data (*RAP2B* and *mir-886*) were also targeted in this experiment. Primer design and the EpiTyper assay were performed by the David H. Murdock Research Institute Genomics core facility according to manufacturer's protocol. Primer sequences are given in Appendix B.3.1.

A second validation experiment was performed for additional DMRs in the 450k data, selected from genes involved in neural development and also including *mir-886*. A total of 20 amplicons in 15 genes associated with neural development were assayed by high-throughput bisulfite sequencing. *mir-886* was the only target included in both validation sets. Samples used for validation included the original AN and TTTS MZ twin pairs, five additional dizygotic (DZ) twin pairs, one set of DZ triplets, parents of all sibling sets, and unaffected siblings as applicable. All sibling sets contained one affected proband, and all family members were unaffected. Methylated and unmethylated human DNAs from Zymo Research Corporation, (Irvine, CA) were also included as controls. Primers were designed using MethPrimer (Li and Dahiya, 2002); in regions

with high CpG density, settings from MethPrimer were used in Primer3Plus (Untergasser et al., 2007) and primers containing a minimum of CpGs selected for use. Primers were ordered from Integrated DNA Technologies (Coralville, IA); sequences and amplification conditions are given in Appendix B.3.2.

Samples were bisulfite converted using the EZ DNA Methylation kit (Zymo), and amplicons generated using HotStarTaq polymerase (Qiagen, Germantown, MD). PCR products were pooled in equal amounts by sample, and sample libraries generated using the Nextera XT sample preparation kit (Illumina, Inc., San Diego, CA). Libraries were sequenced 2x150nt on an Illumina MiSeq. Reads were aligned against target gene excerpts from the hg19 human reference build using Bismark (Krueger and Andrews, 2011) and Bowtie 2 (Langmead and Salzberg, 2012). Due to short amplicon size, overlapping read pairs were merged using FLASH (Magoč and Salzberg, 2011) and all reads aligned as if they had been single-end sequenced. Bismark was used to call methylation at all covered CpGs. Only sites covered by at least 35 reads were included in analyses.

3.1.2 Results

3.1.2.1 Genetic discordance between the twins

A total of 66,231 SNVs passed quality filters, with at least one SNV present in 74% of covered Ensembl genes. The twins had identical genotypes at all sites but one, a

novel mutation in the unaffected twin which was not present in any other unaffected family members. This mutation was located in intron three of sedoheptulokinase (SHPK), a component of the pentose phosphate pathway (Kardon et al., 2008); to our knowledge, neither gene nor pathway has previously been implicated in NTDs. We concluded this sole discordant variant is unlikely to contribute to the twins' different phenotypes.

3.1.2.2 Genetic variants in neural morphology genes

We identified a total of 2,512 variants for these twins in candidate genes for abnormal neural tube morphology (A.1 Genes associated with abnormal neural tube morphology in mice) or exencephaly (A.2 Genes associated with exencephaly in mice). These included 104 nonsynonymous SNVs and one stopgain SNV (in *CASP7*, alternate allele frequency 59.7%). Of nonsynonymous variants, 15 were scored possibly deleterious by both SIFT and either Polyphen-2 HDiv or HVar (Table 14). As described above, genotypes at these sites were common to both twins.

Table 14. Potentially deleterious variants in neural morphology genes, as determined by SIFT and Polyphen-2 scores. AAF, alternate allele frequency; AA, amino acid substitution; GT, genotype; PP2, Polyphen-2; GERP, GERP++ Rejected Substitution score.

SNP	Gene	AAF	AA	SIFT	PP2 HDiv	PP2 HVar	GERP
rs3824915	*ALX4*	---	R35>T	0.01	0.437	0.115	4.61
rs235768	*BMP2*	---	R190>S	0.01	0.978	0.871	5.72
rs16941	*BRCA1*	66.10%	E1038>G	0.04	0.936	0.606	-0.974

rs11074359	*COQ7*	96.10%	T65>M	0.02	0.889	0.631	5.91
rs1110719	*ESPL1*	---	K1435>M	0.02	0.761	0.275	3.65
rs1567047	*FAT4*	---	G3524>D	0.02	1	1	5.77
rs147788385	*FZD6*	---	A150>T	0.03	0.99	0.763	5.6
rs114198341	*HES3*	1.10%	N10>S	0	0.999	0.997	4.51
rs139475791	*HOXB13*	15.10%	R217>C	0	1	0.999	5.31
rs944895	*LAMA5*	99.50%	R3079>W	0.01	0.983	0.545	3.38
rs4667591	*LRP2*	---	I4210>L	0.03	0.995	0.992	4.5
rs45478192	*PALB2*	---	L939>W	0	1	1	5.81
rs139272770	*SHROOM3*	---	H207>Y	0	0.999	0.986	6.06
rs2274924	*TRPM6*	---	K1579>E	0.03	0.787	0.526	4.64
rs1432273	*TTC21B*	88.30%	V201>M	0.03	0.992	0.95	2.52

Using Pearson's chi-squared test, we tested the distributions of exonic vs. intronic and non-synonymous (including stopgain and stoploss) vs. synonymous variants in candidate gene sets as compared to the twins' whole-exome background (Table 15). In both abnormal neural tube morphology and exencephaly candidate genes, we found significant enrichment for exonic SNVs over intronic relative to the rest of the exome, and also for synonymous SNVs over nonsynonymous. These synonymous variants may include alleles with regulatory impact which provided a permissive background for NTDs in these twins and sensitized to environmental risk factors.

Table 15. Enrichments of variant types in candidate gene sets over whole-exome background.

	Exonic vs. Intronic		Nonsynonymous vs Synonymous	
Gene set	*p*-value	Enriched	*p*-value	Enriched
NT morphology	4.31e-04	exonic	1.573e-07	synonymous

Exencephaly	0.0091	exonic	5.69e-04	synonymous
Combined	4.71e-05	exonic	1.89e-07	synonymous

3.1.2.3 Agreement of methylation profiling methods

In terms of the twins' global methylation profile, the MDB-seq and 450k array data agreed favorably with one another (Table 16). Both methods revealed a global tendency towards hypomethylation in the affected twin, and interestingly documented similar numbers of DM sites and associated genes, although very few individual genes were considered differentially methylated by both profiling methods simultaneously.

Table 16. Global overview of methylation as determined by each profiling method.

	MBD-seq	450k array
Total DM regions or sites	1,737	1,152
Hypermethylated sites	45%	34%
Hypomethylated sites	55%	66%
Total DM genes	818	805
Hypermethylated genes	36%	37%
Hypomethylated genes	67%	63%
Genes with hyper & hypo sites	3%	1%

MBD-seq reads covered only 36.6% of all sites queried by the 450k array, including 58.6% of 450k DM sites in this twin pair. Of those, 62.4% showed a difference in MBD-seq read depth which concurred with the direction of change in the 450k data. However, virtually none of these coincided with a statistically significant MBD-seq

DMR, largely due to low read depths in the sequencing data. While greater peaks often occurred just upstream or downstream of the probed CpG, coverage in those areas was not typically deep enough to reliably detect the 13.4% average difference among 450k DM sites. Thus, although global concordance was high, the two datasets did not validate one another at the level of individual DMRs. We therefore focused on the quantitative 450k data for further analysis.

3.1.2.4 Summary of differentially methylated sites from the HumanMethylation450 array

After exclusions based on missing data, detection *p*-value, and predicted non-specific binding, a total of 426,172 sites were included in analysis. A total of 1,153 sites had a difference in methylation of at least 10%, with only 73 surpassing 20% difference. The single greatest delta (+85%) was observed for a CpG in *RAP2B* (chr3:152879313), while the only contiguous region with multiple DM probes was associated with *mir-886* (5:135416205-135416529).

Differentially methylated sites were enriched for GO Biological Process terms relating to cell adhesion, nervous system development, cell communication and signaling, localization, ion transport and homeostasis, and a number of broad high-level terms such as "cell development" and "cellular process" (data not shown). Adhesion term enrichments were largely driven by differential methylation of the protocadherin

locus. Notably, "nervous system development" was the most significantly enriched term ($p=2.68 \times 10^{-9}$) which was neither adhesion-related nor a general global category.

Interestingly, both hypomethylated and hypermethylated genes were enriched for promoter binding motifs of potentially relevant transcription factors (TFs) (Table 17), including the specificity protein family (SP); myeloid zinc finger gene 1 (MZF1), which is an effector of retinoic acid signaling; the Vitamin D receptor (VDR); and Leukemia/lymphoma-related factor (LRF/ZBTB7a). More strikingly, hypermethylated genes showed enrichment for binding motifs of the glucocorticoid receptor (GR), AP-2α and AP-2γ, and CDX1/2. To further investigate the relationship between these TFs and differential methylation in the twins, DM sites were annotated against publicly available ChIP-seq data for AP-2α and CDX2. In total, 22% of DM sites overlapped with at least one TF peak, 20% with CDX2 and 3% with AP-2α.

Table 17. Summary of term enrichments among DM sites by direction of methylation change.

	Hypomethylated	Hypermethylated
Total genes	509 genes	306 genes
GO BP terms	25 terms	27 terms
General GO BP categories	cellular process, cell adhesion, organismal development, organismal signaling, nervous system development, synapse organization	cellular process, localization, transport, organismal signaling, cell surface signaling, ion transport, response to stimulus

promoter TF binding motifs	8 motifs, 6 TFs or families (GATA, LRF, MZF1, SP, TFIIA, VDR)	15 motifs, 11 TFs or families (AP-2, CDX, E2F, GR, HMGIY, LRF, MAF, MZF1, p53, SP, VDR)
3' UTR miRNA binding motifs	1 motif (miR-551b)	2 motifs (miR-139a-3p, miR-560)

3.1.2.5 Contrast of AN twins methylation profile with TTTS twins methylation profile

An additional pair of monozygotic twins with twin-to-twin transfusion syndrome (TTTS) was also run on the array (unpublished data). The affected twin of this pair was anencephalic, which is not typical of TTTS (Sekiya and Hafez, 1977), but could have been secondary to stunted growth and nutrient deficiencies arising from TTTS. We compared the differential methylation profile of these twins to that of the AN twins. Notably, the TTTS twins exhibited half as many DM sites (593) and approximately equal numbers of hypomethylated and hypermethylated loci (49% vs. 51%). In addition, only 32 individual DM CpGs overlapped between these twin pairs, only 12 of which were affected in the same directions. The distinctly different profiles of these twin pairs suggest that each pair's methylation differences are not likely attributable to site-specific variance, but reflect the separate etiologies underlying each case.

3.1.2.6 Validation of selected differentially methylated sites

Both validation methods, the EpiTyper assay and targeted resequencing, faced considerable difficulties in amplicon generation. On average, about 40% of samples dropped out at any given site by either method, though the actual proportion of missing data varied widely. Data was obtained for both members of the original MZ twin pair at six EpiTyper target regions (85%) and nine resequencing target regions (45%), including three ampicons tiling the *mir-886* locus. Roughly half of covered regions validated as differentially methylated between these twins (Table 18). On average, percent methylation as quantitated by resequencing was 10% lower than the 450k value for the affected twin, and 15% lower in the unaffected. An additional five resequencing targets (25%) were covered in just one of the original pair. At three of these sites (in *APC2*, *KIRREL3*, and *MED12*), methylation values in the covered twin were comparable to those obtained with the 450k array (data not shown).

Table 18. DM loci that validated in the original MZ twin pair. Delta values compare the affected to the unaffected.

Locus	DMR	Original	Delta	Validation	Delta
LRP1B_CpG_4	2:142313484-142313683	MBD-seq	hypo	EpiTyper	-27%
APC_CpG_1	5:112116370-112116577	MBD-seq	hypo	EpiTyper	-17%
APC_CpG_2	5:112116370-112116577	MBD-seq	hypo	EpiTyper	-26%
mir-886	5:135416205-135416029	450k	-25% (avg)	EpiTyper MiSeq	-8% (avg) -10% (avg)

NR1H3	11:47279365	450k	+10%	MiSeq	+10%
SPTBN4	19:40995921	450k	+11%	MiSeq	+16%

Three EpiTyper target regions (43%) and five resequencing target regions (20%) were covered in both MZ twins and did not validate the originally observed differential methylation (Table 19). Most of these loci returned as either fully methylated or fully unmethylated in the validation experiment, with the exception of the site in *TEAD1*. Unmethylated controls for *CAMTA1, DAAM2*, and *TEAD1* exhibited greater methylation than expected, suggesting possible systematic issues with these target regions.

Table 19. DM sites that did not validate in the original MZ twin pair.

Gene	DMR	Original	Delta	Validation	Values (A, U)
CAMTA1	chr1:6871637-6871846	MBD-seq	hyper	EpiTyper	93%, 91%
DAAM2	6:39784553-39784752	MBD-seq	hypo	EpiTyper	99%, 96%
RAP2B	chr3:152879313	450k	+85%	EpiTyper	2%, 4%
GRM8	7: 126893362	450k	+12%	MiSeq	0%, 0%
HES7	17: 8027961	450k	-26%	MiSeq	1%, 1%
SMPD3	16: 68482518	450k	-11%	MiSeq	1%, 0%
TEAD1	11: 12708608	450k	-13%	MiSeq	37%, 30%
ZBTB16	11: 114068801	450k	+17%	MiSeq	93%, 96%

3.1.2.7 Differential methylation of *mir-886*

As the only set of contiguous DM probes identified from the 450k array, we included *mir-886* in both EpiTyper and targeted resequencing validation experiments.

Despite the high density of associated CpGs, *mir-886* was not covered by the MBD-seq assay, likely because both of the twins had low methylation across this region. This DMR was the most robustly validated, with all associated CpGs being about 10% less methylated in the affected twin. By resequencing, unaffected family members showed "normal" methylation of near 50% across the region (Figure 7A).

Among DZ twin pairs, we found that most had normal methylation across the region (Figure 7B). Two affecteds and three unaffecteds, all of different families showed very low methylation (<10%) across the region, as did two parents and an unaffected sibling of DZ twins (Figure 7C). The remainder of individuals had methylation levels in the range of about 35% to 65%, with females generally slightly less methylated than males (Figure 7C).

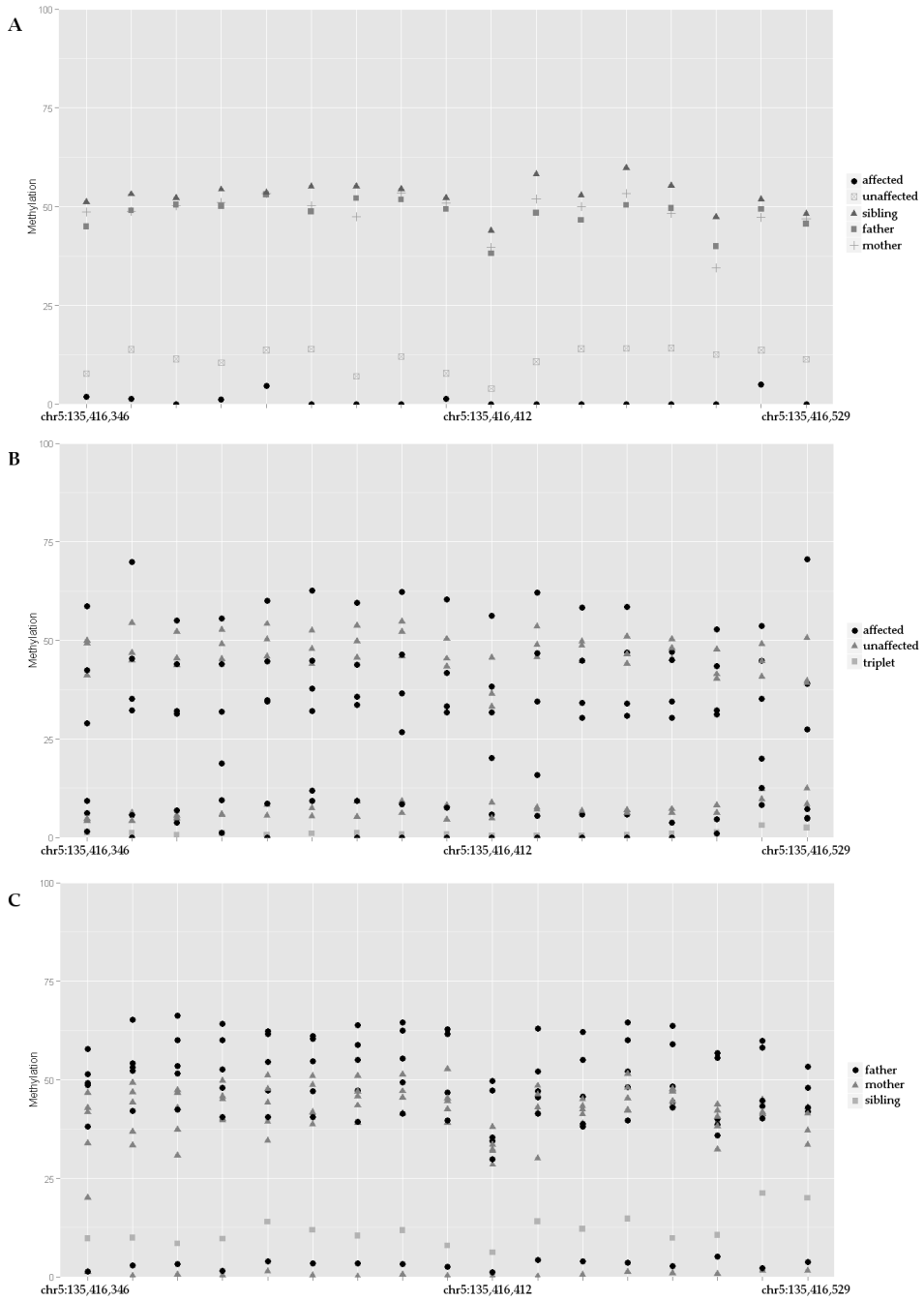

Figure 7. Plots of methylation across the mir-886 locus. A) MZ twins, their parents, and unaffected sibling; B) DZ twins; C) family members of DZ twins.

3.1.3 Discussion

In this monozygotic twin pair discordant for anencephaly, we identified no SNVs potentially causative for their divergent phenotypes. Candidate genes for both abnormal neural tube morphology and exencephaly exhibited significant enrichments relative to the rest of the exome for exonic SNVs over intronic and for synonymous SNVs over nonsynonymous (Table 15). This overrepresentation of synonymous variants may include alleles with regulatory impact that provided a permissive background for NTDs in these twins and sensitized to environmental risk factors.

We did identify distinct differences in the twins' epigenetic profiles (Table 16), corroborated by two different profiling methods. Notably, hypomethylated genes were enriched for members of the GO Biological Process term 'nervous system development' (GO:0007399), while hypomethylated genes were enriched for promoter binding motifs of the NTD candidate transcription factors CDX1/2 and AP-2α. Using publicly available ChIP-seq data, we found that 22% of DM sites overlapped with footprints of the CDX2 transcription factor. These findings suggest that epigenetic misregulation at binding sites for NTD-related transcription factors may contribute to NTD etiology. In the context of this twin pair, one or more such epigenetic events may have acted as a "second hit" in the affected twin, leading to the development of anencephaly.

Overall, MZ twins are much more similar in epigenetic profile than are siblings or DZ twins. A study of 80 MZ twin pairs found that 65% of twins had near-identical genomic methylation and histone levels; the ages of twins included in this study ranged from 3 to 74 years, with younger twins having higher concordances (Fraga et al., 2005). However, even neonatal MZ twins do not have truly identical DNA methylation profiles (Ollikainen et al., 2010). These observed differences, given the near-identity of MZ twin genomic sequences, speak to the impacts of environmental and stochastic factors on DNA methylation. We additionally profiled a twin pair with TTTS syndrome on the 450k array and found distinct differences between them and the AN twins; namely, half as many sites were differentially methylated, and the TTTS twins had no skew towards hypomethylation. Furthermore, only 12 individual CpGs were differentially methylated in the same direction in both twin pairs.

However, we found that only about half of the DM sites covered by our validation experiments did in fact validate, whether by EpiTyper assay or by targeted resequencing. Some of this discrepancy may be attributable to the PCR process as both validation methods were amplicon-based, however neither of our initial assays relied on PCR. Additionally, we noticed in the resequencing experiment that the affected MZ twin generally quantitated 10% lower than it had on the 450k assay, while the unaffected twin quantitated on average 15% lower. As our average 450k DM CpG had a magnitude

change of only 13%, this sample-specific shift may have masked some true differences in the validation, particularly for hypermethylated sites. Finally, it has been calculated that the Infinium probes can detect between 20% to 7% differential methylation at the 95% confidence level, with greatest sensitivity at high or low methylation values (Bibikova et al., 2009). Our choice of 10% delta as the threshold for differential methylation, combined with a very small sample size, likely permitted a greater number of false positives among sites in the middle range.

Our single most robust finding is differential methylation across the *mir-886* locus (Figure 7). Both of the MZ twins had low methylation across this region, but the affected was about 10% less methylated than the unaffected by all three assays. Among DZ twins, two affecteds had similarly low methylation values, and two others were on the low end (<35%) of what appears to be normal (35-65%). Conversely, six unaffecteds, whether twins or family members, had very low methylation, while 14 fell in the apparent normal range. This suggests a tendency for NTD cases to be hypomethylated in this region, although hypomethylation is clearly not sufficient to cause NTDs in and of itself. Additionally, we observed that mothers in these families were generally less methylated than fathers, an intriguing finding in the light of the female excess documented for anencephaly cases.

What significance hypomethylation at this locus may have is difficult to determine. The gene itself is a subject of some controversy; initially identified as a microRNA, it has since been reclassified as a vault RNA based on sequence similarity. However, that similarity is markedly less than the conservation which exists between other vault RNAs, and pre-mir-886 does not colocalize with vault proteins (Lee et al., 2011). At the same time, it is inefficiently processed into two mature miRNAs by Dicer, suggesting that this ncRNA may fall into another class entirely. Of its derivative miRNAs, mir-886-5p has been shown to be upregulated in brain tissue of anencephaly cases (Zhang et al., 2010), consistent with potential hypomethylation of the locus. Mir-886-5p has also been documented as regulating *BAX*, a pro-apoptotic factor (Li et al., 2011b), while mir-886-3p may regulate cell proliferation and migration (Xiong et al., 2011); the balance of these processes is critical to successful neural tube closure. Recent literature suggests maternal imprinting of *mir-886*, which may be the first polymorphic imprinted locus identified in humans (Romanelli et al., 2014). In any case, *mir-886* is clearly an interesting gene that merits further investigation.

One drawback of our study is the deficit of neural tissue for epigenetic analysis. First and foremost, neurulation happens over days 26-30 of human gestation, for which period very few embryos are available to be collected. Human studies must, by necessity, use more-differentiated tissues obtained later in gestation. However,

obtaining neural tissue from anencephaly cases even at the end of term is difficult because much of the brain degrades during gestation; furthermore, matching samples cannot be obtained from their unaffected, viable twins. Conversely, cord blood is readily obtainable for both anencephalic and normal twins.

Though the DNA used in this experiment came from PBMCs, and the tissue-specific aspects of epigenomic profiles have been soundly attested, we believe our findings are nonetheless informative. First, while much research has focused on tissue-specific methylation, high correlations ($r = 0.859$) have been documented of methylation state between multiple tissues (Byun et al., 2009). Recent research has also demonstrated that between-individual differences in methylation are significantly correlated between brain tissues and blood, indicating that any given change in methylation may be detectable in both tissues (Davies et al., 2012). Consistent with this correlation, several studies have identified loci in blood which are differentially methylated in the context of neurological diseases (Gregory et al., 2009; Uddin et al., 2010). In all, this evidence supports that the differences we observed in these twins are likely to have been present in more proximal tissues as well.

It is possible that the discordant MZ twins, although monozygotic, had some genetic variation we did not identify. Although exome capture is considered to query the "entire exome", in practice a number of regions (esp. first exons) are not sequenced.

By definition, exome sequencing also omits many intronic, UTR, and intergenic sites at which variants could have regulatory effects. The MBD-seq approach we used in this study also has several limitations of note. Although more sensitive than methylcytosine immunoprecipitation, MBD-seq does not effectively capture sequences containing few CpGs or sites with very low methylation (Li et al., 2010). Such sites could be functionally relevant but not represented in our data. Additionally, this method cannot identify individual methylated bases, nor provide percent methylation of a site, and its sensitivity is dependent upon read depth. Conversely, the 450k array provides quantitative methylation values at single-base resolution, but is limited to only those sites represented on the array. Used in combination, we found these methods to corroborate very well at the genomic level, and with sufficient sequence read depth, they may also corroborate at the site level.

In this experiment we were not able to evaluate expression levels for genes associated with DMRs, another important piece of functional evidence. However, this is to our knowledge the first study to investigate differential methylation in NTD cases at the level of genes and individual CpGs, and so represents an important step forward in understanding the role of altered methylation in NTD etiology.

Chapter 4: Interactions with the environment: the case of fumonisin

With a myriad of genetic and environmental factors potentially contributing to NTD etiology, it stands to reason that gene-environment interactions may also be substantial determiners of NTD risk. Here we examine the confluence of genetic differences and/or epigenetic outcomes of fumonisin exposure both in a mouse model system and in healthy women from Guatemalan communities.

4.1 Differential methylation in LM/Bc and SWV mouse embryos exposed to FB1

Despite a partially shared genetic background, LM/Bc and SWV mouse strains have strikingly different susceptibilities to fumonisin-induced NTDs (Section 1.5). The immediate effect of fumonisin is competitive inhibition of ceramide synthases (CERS1 and related proteins), elevating levels of simple sphingolipids from which ceramide would otherwise be made, and reducing levels of more complex ceramide derivatives. Maternal supplementation of folate or ganglioside GM1, a ceramide-derived component of lipid rafts that may anchor folate receptors, has been shown to rescue FB1-induced NTDs in LM/Bc embryos (Gelineau-van Waes et al., 2005). However, the mechanism by which FB1 exposure causes NTDs and the aspect of folate biology that rescues the phenotype remains unknown.

Both ceramide and the simple sphingolipids, sphingosine (So) and sphinganine (Sa), are bioactive molecules that act as secondary messengers which contribute to the regulation of many processes (Section 1.4.2.2). In addition, simple sphingolipids are known to directly inhibit histone deacetylases (Hait et al., 2009), indicating a role for these molecules in epigenetic regulation. The Gelineau-van Waes lab has previously shown that sphinganine-1-phosphate (Sa1P) in particular is highly upregulated in LM/Bc mice exposed to FB1 (Gelineau-van Waes et al., 2012), and that histone acetylation is greatly increased in LM/Bc embryos with FB1-induced NTDs (unpublished data). This evidence suggests a role for epigenetic modifications in mediating the consequences of FB1 exposure, especially as folate is a required cofactor for methylation reactions. To elucidate these mechanisms, we compared the epigenetic profiles of control and FB1-treated embryos of each strain to determine how FB1 exposure impacts the epigenome, and whether those epigenetic changes may contribute to NTD etiology in the LM/Bc strain.

4.1.1 Methods

4.1.1.1 Experimental animals, treatment, and sample collection

Inbred SWV and LM/Bc mice were maintained by the Gelineau-van Waes lab at Creighton University as previously described (Gelineau-van Waes et al., 2012). All animal procedures were performed in accordance with the Public Health Service (PHS)

policy on Humane Care and Use of Laboratory Animals, and approved by the Creighton University Institutional Animal Care and Use Committee. Fumonisin B1 was purchased from PROMEC (Tygerberg, South Africa), and administered to pregnant SWV and LM/Bc dams on E7.5, E8.5, and E9.5 via intraperitoneal injection (20 mg/kg/day) in a volume of 0.1 ml per 10 gm body weight. Pregnant animals in the control groups were given saline vehicle at the same time points via intraperitoneal injection (0.1 ml per 10 mg body weight). Embryos were collected four hours after the final injection, and DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI), scaled down according to tissue input amounts.

4.1.1.2 Methyl-seq library preparation and sequencing

Regulatory regions were enriched and libraries prepared for bisulfite sequencing using the SureSelect Mouse Methyl-seq XT kit (Agilent Technologies, Santa Clara, CA) according to manufacturer's protocol. This kit covers 109 Mb of the mouse mm9 genome and is designed to target CpG islands, shores, shelves, Ensembl regulatory features, and Open Regulatory Annotation features. Briefly, extracted DNA was sheared on a Covaris S220 Focused-ultrasonicator and target regions were enriched through hybridization with and capture of biotinylated complementary RNA probes. Fully methylated adapters were ligated to enriched fragments, then samples were bisulfite converted using the EZ DNA Methylation-Gold kit (Zymo Research, Irvine,

CA) and indexed for sequencing. Libraries were pooled two per lane and sequenced 100 nt paired-end on an Illumina HiSeq instrument.

4.1.1.3 Methyl-seq data analysis

Reads were aligned to the mouse mm9 reference using Bismark (Krueger and Andrews, 2011) and Bowtie 2 (Langmead and Salzberg, 2012). Bismark was also used to remove duplicates from the alignments. The program Bis-SNP (Liu et al., 2012), built upon the GATK framework but specific for bisulfite-converted sequence analysis, was used for local refinement around indels and base quality score recalibration. Bis-SNP was also used for methylation calling at every CpG within 400 nt of probe binding sites. Only sites with at least 10 reads in all four samples were included in downstream analyses. Sites were annotated against RefSeq genes using ANNOVAR (Wang et al., 2010) with a near gene threshold of 2 kb. Methylation values, a 'T' at a CpG dinucleotide being considered un-methylated and a 'C' considered methylated, were compared between strain controls, between treated and control groups, and between the LM/Bc FB1-treated embryo and all three other embryos. At least 20% difference in methylation at a base pair specific coordinate between 'T' and 'C' reads was required to consider a site differentially methylated. Regions were considered differentially methylated between samples if they contained two or more CpGs within 200 nt which all differed by 20% or more in the same direction.

4.1.1.4 Validation of DM sites by high-throughput bisulfite sequencing

Twenty DMRs localized to genes in sphingolipid metabolism, Wnt signaling, or Hippo signaling were selected for validation by targeted bisulfite resequencing. Primer design, PCR amplification, sequencing, and analysis were all performed as previously described (Section 3.1.1.5). A total of 43 samples, plus an enzymatically methylated mouse control (Zymo Research), were used as the validation cohort (Table 20). A few whole-embryo samples were included to examine the consistency of methylation between head tissue and entire embryo. Significance of methylation changes was evaluated for head samples only by two-way ANOVA incorporating strain, treatment, and strain x treatment interaction; p-values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) correction.

Table 20. Mouse sample cohort for validation by targeted resequencing.

	LM/Bc		SWV	
	control	FB1	control	B1
Whole embryo	2	1	2	2
Embryo head	8	9	9	10

4.1.2 Results

4.1.2.1 Overview of methyl-seq coverage and alignment quality

All samples showed good alignment to reference and low duplications rates (Table 21). Samples were very similar in base quality per cycle, read quality, sequence

content, and GC content as reported by FastQC (data not shown). Additionally, all four samples had nearly identical aggregate methylation levels (38%) across profiled CpGs. A total of 796,195 CpGs were covered by at least 10 reads in all four samples, and included in downstream analyses.

Table 21. Alignment metrics for the mouse methyl-seq data.

	Total Read Pairs	Mapped	Duplicated
LM/Bc ctrl	16777129	70.7%	5%
LM/Bc FB1	17914705	69.0%	13%
SWV ctrl	25782553	70.9%	10%
SWV FB1	31823266	71.0%	6%

4.1.2.2 Differential methylation specific to LM/Bc with FB1 exposure

We identified 31,726 differentially methylated sites specific to the LM/Bc FB1-treated embryo. Of these, 6,685 (21%) were within 200 nt of another DM CpG having the same direction of change, comprising 2,921 DMRs. Just under one-third of DM CpGs (2,123 sites) were intergenic, and excluded from further analysis. The remaining 4,562 sites were distributed across 1,965 genes, averaging 1 DMR (2.3 DM CpGs) per gene. Gene Ontology (GO) Biological Process (BP) term enrichments for these genes were determined using g:Profiler with g:SCS multiple testing correction (Reimand et al., 2011). In all, 387 BP terms were significantly enriched; a best-per-parent summary is given in Table 22.

Many enriched terms were processes with relevance to NTDs, including nervous system development (p=9.01e-23), its child terms forebrain development (p=5.89e-03) and hindbrain development (p=4.92e-02), epithelial tube morphogenesis (p=2.42e-02), actin cytoskeleton organization (p=4.72e-03), cell proliferation (p=5.60e-05), purine nucleotide metabolic process (p=7.39e-04), calcium ion homeostasis (p=1.04e-03), and GTP metabolic process (p=2.22e-02). Collectively, these enrichments suggest that epigenetic regulation of processes contributing to neural development and neural tube closure may be particularly impacted by FB1 exposure and/or ceramide synthase inhibition.

Table 22. Best-per-parent summary of GO Biological Process term enrichments for genes with DMRs in the LM/Bc FB1-treated embryo.

p-value	GO Biological Process Term	Term ID	Genes
1.14e-32	single-organism process	GO:0044699	1235
5.55e-30	developmental process	GO:0032502	563
5.61e-28	multicellular organismal development	GO:0007275	503
4.85e-26	single-organism cellular process	GO:0044763	1108
1.79e-23	cellular component organization	GO:0016043	500
2.55e-23	localization	GO:0051179	529
5.30e-22	positive regulation of biological process	GO:0048518	462
9.87e-20	cellular metabolic process	GO:0044237	835
4.16e-19	intracellular signal transduction	GO:0035556	276
4.62e-17	cell adhesion	GO:0007155	148
1.44e-10	locomotion	GO:0040011	167
2.89e-09	enzyme linked receptor protein signaling pathway	GO:0007167	103
1.48e-04	regulation of GTPase activity	GO:0043087	57
2.69e-04	cellular response to organic substance	GO:0071310	153
2.96e-04	growth	GO:0040007	108
4.28e-03	muscle contraction	GO:0006936	37

4.1.2.3 Overview of targeted resequencing results

Of the 20 validation amplicons, all but one were sequenced to a depth of 35x or more across at least 35 samples (80%). The *Tead4* amplicon was covered to 35x in only 30 samples (68%). Data was obtained for a total of 134 CpGs, though due to the tagmentation method of library preparation, CpGs near amplicon ends were sequenced less deeply. The fully-methylated control was covered at 107 CpG sites, and had greater than 85% methylation for all but seven. Those seven were excluded from further analyses. Generally, methylation for whole-embryo samples was similar in magnitude, but did appear to group separately from head tissue counterparts (data not shown). Though we did not have enough embryo samples to test significance, this suggests that methylation profiles of different tissues may be distinct even in early development.

In total, 78 CpGs (58%) were found to have at least nominally significant association of methylation level with strain, treatment group, or both (Table 23). A majority of strain associations (63%) passed multiple testing correction, while few did in other models, likely due to the small sample size and sometimes considerable within-strain variability of methylation (see below). Only three amplicons (Ctnnd2_B, Fat3, and Vgll4_A) exhibited no significant associations under any model.

Table 23. Summary of amplicons with significant associations (p<0.05) between methylation, strain, and treatment group.

		Strain x Treatment		Strain		Treatment	
Amplicon	CpGs	Nominal	FDR	Nominal	FDR	Nominal	FDR
Ctnnd2_A	5	1	---	---	---	2	---
Ctnnd2_B	3	---	---	---	---	---	---
E2f2	6	2	---	---	---	---	---
Fat3	3	---	---	---	---	---	---
Frmd4a	3	---	---	---	---	1	---
L3mbtl1_A	6	---	---	1	1	---	---
L3mbtl1_B	10	6	---	9	9	1	---
L3mbtl1_C	10	1	---	10	7	6	---
Nf2	7	---	---	1	---	---	---
Prickle1_A	6	---	---	1	---	1	---
Prickle1_B	5	1	---	2	---	1	---
Shmt2_A	8	---	---	3	1	---	---
Shmt2_B	10	---	---	---	---	1	---
Smad6	3	3	2	---	---	---	---
Tead4	12	---	---	3	1	1	---
Vgll3	9	---	---	3	---	---	---
Vgll4_A	8	---	---	---	---	---	---
Vgll4_B	5	4	---	---	---	---	---
Wwc2_A	8	2	---	1	---	1	---
Wwc2_B	7	2	---	7	7	---	---

4.1.2.4 Strain-specific changes in methylation associated with FB1 exposure

Of 20 amplicons in the validation set, nine (45%) were found to contain at least one CpG with a significant, strain-specific change in methylation with FB1 exposure. The most significant SNPs per amplicon are shown in Figure 8. Generally, the

differences in methylation are small (<15%) both between strains and between treatments, consistent with other literature on non-cancer-related methylation changes (Dayeh et al., 2014; Rakyan et al., 2011).

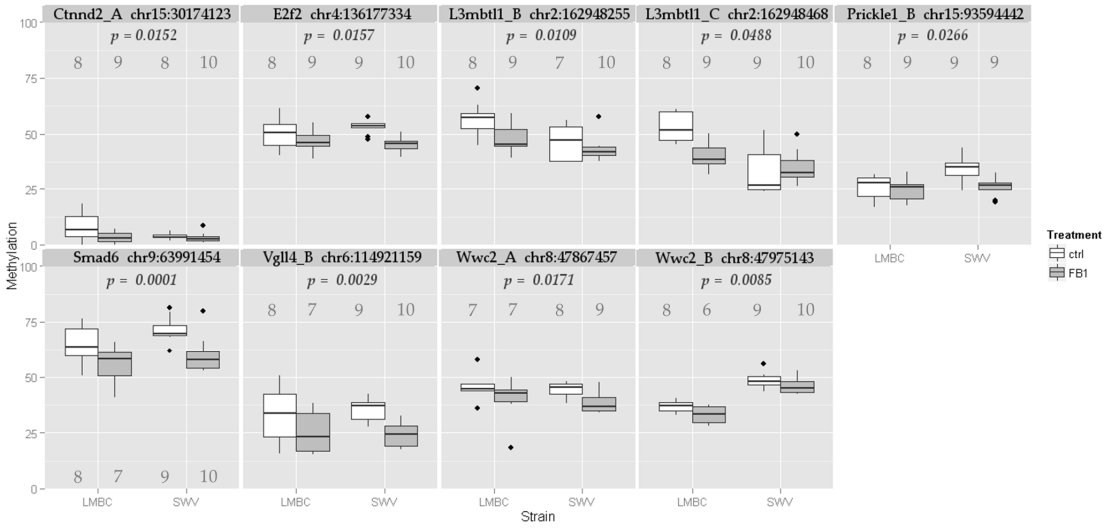

Figure 8. Methylation vs. treatment group for CpGs with significantly different responses between strains upon FB1 exposure. Numbers on plots indicate uncorrected *p*-values and *n* per treatment group.

Four additional amplicons contained sites which did not quite achieve nominal significance: Frmd4a, Shmt2_A, Shmt2_B, and Tead4 (Figure 9). Of these, two show apparently greater effect in one strain (Frmd4a, Tead4) or effects only in LM/Bc (Shmt2_A, Shmt2_B).

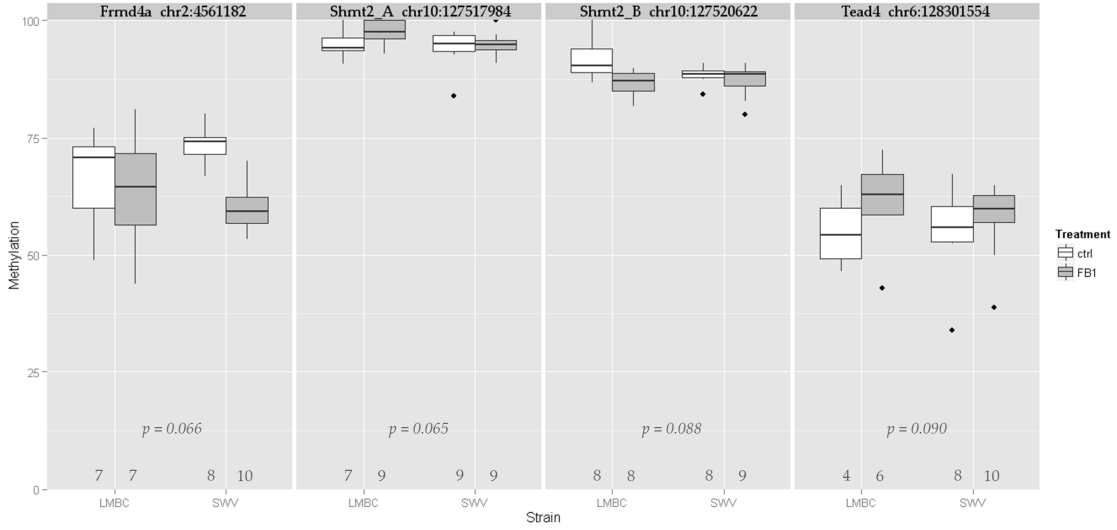

Figure 9. Sites from amplicons with suggestive trends in methylation. Numbers on plots indicate uncorrected *p*-values and *n* per treatment group.

4.1.3 Discussion

We identified 4,562 DMRs in the FB1-treated LM/Bc embryo compared to strain controls and a FB1-treated SWV embryo. Genes with DMRs were significantly enriched for a number of GO BP terms relating to neural development and neural tube closure, including nervous system development (p=9.01e-23), actin cytoskeleton reorganization (p=4.72e-03), cell proliferation (p=5.60e-05), purine nucleotide metabolic process (p=7.39e-04), and others. These findings suggest that epigenetic regulation of genes involved in neural development may be particularly impacted by FB1 exposure.

Of 20 selected DMRs from sphingolipid metabolism, Wnt signaling, and Hippo signaling genes, nine (45%) validated as having significant, strain-specific differences in methylation for at least one CpG (Table 23). Four others approached significance

(Figure 9), suggesting a strain-specific trend in response at these loci. In all, 65% of sites selected for validation appear to have at least a trend towards strain-specific epigenetic changes with FB1 exposure. This is similar to the validation rate obtained for our human methylation study (Section 3.1.2.6), and was likely similarly affected by the small sample size of our initial experiment (n=1 each strain and treatment).

Some DMRs with strain-specific responses showed a strong change only in one strain (e.g. *Fat3, Tead4, Wwc2*), but at other sites (chiefly in *L3mbtl1*), LM/Bc and SWV mice were affected in opposite manners (Figure 8). These data indicate that genetic background can influence both magnitude and direction of epigenetic change.

In this study, we did not investigate gene expression differences. The Gelineau-van Waes lab has previously run Affymetrix gene expression arrays comparing LM/Bc FB1-treated and control embryos (data not yet published). Of all genes with DMRs, 176 (9%) were significantly differentially expressed, including five in the validation set (*Ctnnd2*, *E2f2*, *Frmd4a, Shmt2,* and *Vgll3*). These genes are strong candidates for epigenetic regulation at their associated differentially methylated loci. We do not currently have data on gene expression differences between LM/Bc and SWV strains, but these experiments are in progress.

4.2 Genetic variants influencing the effects of fumonisin on sphingolipid metabolism in Guatemalan populations

Our mouse model of FB1-induced NTDs supports the notion that genetic susceptibility modulates the metabolic and developmental repercussions of *in utero* fumonisin exposure. While there is no conclusive evidence showing that FB1 causes NTDs in humans, it is likely that similar genetic mechanisms influence the consequences of FB1 exposure in humans. To identify potential modifiers of human FB1 exposure, we performed a SNP association study in participants recruited from Guatemalan departments with generally high (Jutiapa) or low (Escuintla, Chimaltenango) prevalences of FB1 in maize (Torres et al., 2014).

4.2.1 Methods

4.2.1.1 Sample collection and DNA extraction

Participants were recruited by the Centro de Investigaciones en Nutrición y Salud, Guatemala City, Guatemala. The research protocol and consent form were approved by the Comité Institucional de Ética of the Instituto de Nutrición de Centro América y Panamá (Project Number CIE REV003/2010), a U.S. Department of Health, and Human Services Office for Human Research Protections registered Independent Ethics Committee. The Comité Nacional de Ética of the Guatemalan Ministry of Health also gave its ethical approval for this project.

Participants were recruited from villages in the Guatemalan departments of Escuintla, Chimaltenango, and Jutiapa over the course of 2011 and 2012. Blood spots from finger pricks and urine spots were collected for each participant on separate Whatman 903 filter cards. Sphingolipid metabolites (Sa1P and So1P) in blood spots and FB1 in urine were quantified via mass spectrometry by the Riley lab with the USDA – ARS Toxicology and Mycotoxin Research Unit (Torres et al., 2014). DNA was directly whole-genome amplified from blood spot lysates using the REPLI-g Mini Kit (Qiagen, Hilden, Germany), following an adapted protocol available from the manufacturer's website. Amplified DNA was purified using the DNA Clean & Concentrator-5 kit (Zymo Research).

4.2.1.2 GoldenGate genotyping assay

A custom genotyping assay using the GoldenGate Universal-32 BeadChip (Illumina, Inc., San Diego, CA) was designed to target 125 ancestry-informative markers (AIMs) previously characterized for use in American populations (Kosoy et al., 2009) and 259 haplotype-tagging SNPs in sphingolipid metabolism, Wnt signaling, and Hippo signaling genes (Table 24). AIMs were used to exclude probable first-degree relatives from analyses and to evaluate population stratification. Haplotype-tagging SNPs were identified from HapMap release 3 MEX population genotypes using ldSelect (Carlson et al., 2004). Target genes and haplotype bins are summarized in Table 24. One DNA

standard from the Centre d'Etude du Polymorphisme Humain and one duplicated sample were included on each plate as performance controls.

The GoldenGate assay was performed according to manufacturer's protocol by the Duke NTD team. Internal spiked-in controls were used to evaluate chip performance, including measures of intensity and contamination. All chips had generally normal values for internal controls except plate A, which exhibited low intensities. This was most likely caused by a poorly performing thermocycler, which was not used for subsequent plates. Genotype clusters were evaluated separately for plate A.

Table 24. Genes represented on the custom GoldenGate genotyping assay. Number of haplotype bins indicated in parentheses.

CELSR1 (34)	*CERS1-GDF1* (1)	*COL3A1* (4)	*CTNNB1* (5)	*DAAM1* (12)
DVL1 (2)	*DVL2* (2)	*DVL3* (2)	*EDG5* (2)	*GDF3* (3)
LPP3 (17)	*LRP6* (6)	*PAX6* (5)	*S1PR3* (5)	*SGPL1* (10)
SGPP1 (2)	*SGPP2* (18)	*SHMT1* (3)	*SMAD2* (5)	*SPHK1* (2)
SPHK2 (1)	*TCF7L1* (31)	*TCF7L2* (38)	*TEAD1* (24)	*VANGL1* (19)
VANGL2 (4)	*VGLL3* (2)			

4.2.1.3 Analysis of genotype data

Genotyping performance and statistical analyses were performed by the Duke NTD team. As only 384 SNPs were included on the assay, genotype clusters were individually evaluated for amplification success, cluster separation, and number of

apparent clusters. SNPs failing these criteria were excluded. Samples with a call rate of less than 90% were excluded from analyses. Identity-by-descent (IBD) estimation was performed in PLINK (Purcell et al., 2007) using all SNPs to remove duplicate individuals. Ancestry estimation was performed on ancestry-informative markers in STRUCTURE (Pritchard et al., 2000); apparent first-degree relatives and individuals with uncertain family membership were excluded. These markers were also evaluated for population stratification using eigenstrat (Price et al., 2006). An additional seven samples were removed for mismatches with earlier genotype data. Association of SNPs with blood So1P or Sa1P level was evaluated using linear regression models in PLINK with age, location, urine fumonisin level, and population structure as covariates.

4.2.2 Results

4.2.2.1 Population stratification and pruning of probable close relatives

Upon clustering review, 63 SNPs were removed for poor performance. A total of 321 SNPs were included in analyses, including 105 AIMs and 216 tag SNPs. Population stratification analysis revealed that participants were more similar to MEX and CEU populations than Asian (Figure 10A) or YRI (data not shown). Samples generally clustered with the MEX population (Figure 10B). A number of samples fell just outside the MEX cluster, but did not clearly separate; this group may include individuals with

greater Mayan ancestry. In general, however, these data support our use of the MEX population data for identifying haplotype tagging SNPs in this study.

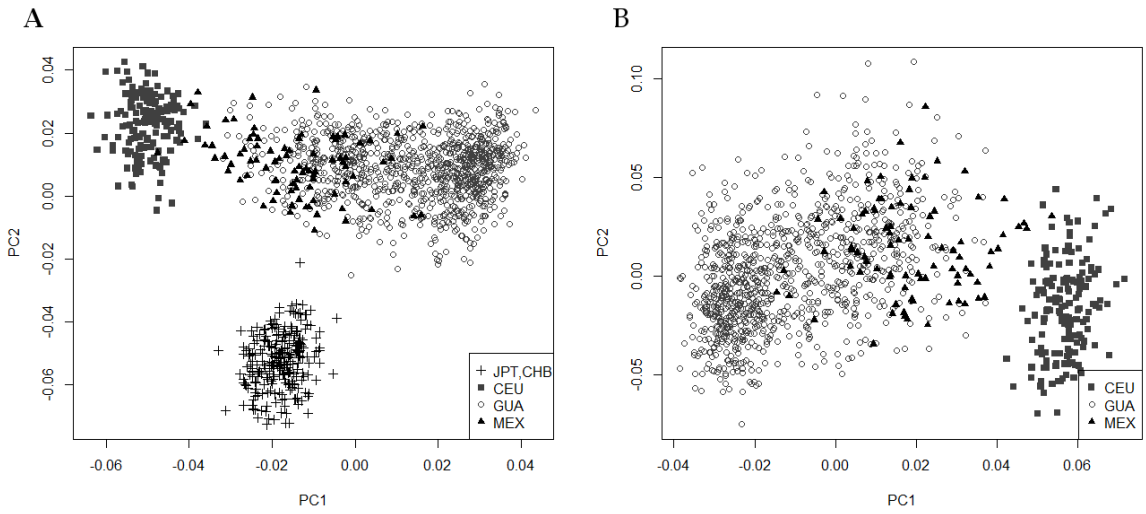

Figure 10. Principal component plots showing population stratification of samples. A) with HapMap Asian population; B) with CEU and MEX populations only.

Ancestry estimation identified 82 definite first-degree relatives. For each family group, only the sample with highest call rate was retained. Another 206 samples were excluded for relatedness where no clear family structure could be determined. An additional 19 individuals were found to have duplicate samples based on IBD estimation. Ultimately, a total of 440 samples were included in association analyses.

4.2.2.2 Association of variants with phenotypic outcomes of fumonisin exposure

Of 216 haplotype-tagging SNPs included in analysis, 22 were found to have nominally significant associations with sphingolipid metabolite levels (Table 25), 13 with So1P and 20 with Sa1P. Eleven variants were significantly associated with both

sphingolipids. Only one association, that of rs10765992 with Sa1P level, remained significant after Bonferroni correction. Interestingly, most genes are components or targets of Wnt signaling; the only sphingolipid metabolism genes containing associated SNPs were *PPAP2B* and *SGPL1*.

Table 25. Variants significantly associated with sphingolipid metabolite levels. All p-values are uncorrected.

SNP	Gene	Class	*p* (Sa1P)	*p* (So1P)
rs10765992	*TEAD1*	intronic	0.0001	0.00172
rs10500759	*TEAD1*	intronic	0.0006	0.00458
rs5767230	*CELSR1*	intronic	0.0070	0.00070
rs644242	*PAX6*	intronic	0.0023	0.00820
rs3026390	*PAX6*	intronic	0.0311	0.00288
rs2071754	*PAX6*	intronic	0.0223	0.00486
rs4757953	*TEAD1*	intronic	0.0061	ns
rs1122940	*CELSR1*	intronic	0.0080	0.01966
rs10885398	*TCF7L2*	intronic	0.0135	0.03067
rs1866714	*TEAD1*	intronic	0.0136	ns
rs11022509	*TEAD1*	intronic	0.0171	ns
rs2296783	*TCF7L2*	intronic	0.0181	0.04817
rs6741339	*TCF7L1*	intronic	ns	0.02048
rs1516454	*COL3A1*	upstream	0.0208	0.03420
rs17096088	*DAAM1*	intronic	ns	0.02751
rs8139582	*CELSR1*	intronic	0.0287	ns
rs17411627	*TEAD1*	intronic	0.0311	ns
rs7513382	*VANGL2*	intronic	0.0335	ns
rs1984519	*CELSR1*	intronic	0.0404	ns
rs3788723	*CELSR1*	intronic	0.0408	0.04904
rs2065553	*PPAP2B*	intronic	0.0453	ns
rs10823627	*SGPL1*	intronic	0.0471	ns

4.2.3 Discussion

We identified 22 variants in ten genes that were significantly associated with sphingolipid levels in blood after fumonisin exposure (Table 25). Eleven of these were significantly associated with both So1P and Sa1P levels. Only one association survived multiple testing correction, that of rs10765992 in *TEAD1* with Sa1P level. The association of *TEAD1* variants with sphingolipid levels in humans is particularly interesting, as this is an NTD candidate gene for which we also identified genetic variants between LM/Bc and SWV mice (Section 2.2.2.5) and a novel variant in our multiplex AN family (Section 2.1.2.5).

Interestingly, only two genes with significant associations, *SGPL1* and *PPAP2B*, were directly involved in sphingolipid metabolism. The remainder was largely comprised of genes involved in Wnt signaling, suggesting the Wnt pathway may play an important role in regulating blood sphingolipid levels. Sphingolipids have been shown to regulate Wnt and β-catenin signaling (Pederson et al., 2008; Schmelz et al., 2001), but to our knowledge this is the first indication that Wnt genes also regulate sphingolipid metabolism.

4.3 Conclusions

The intersections of genetic variation, environmental factors, and the epigenome in regulating neural tube closure only underscore how complex this process is. We show

here that environmental factors may affect the epigenome in ways which could have detrimental impacts on gene expression (Section 4.1.2.2). We have also identified haplotype-tagging variants which associate with phenotypic outcomes of FB1 exposure in humans (Sections 2.2.2.2, 4.2.2.2). The functional impacts of these haplotypes remain to be determined, but taken together, these findings suggest a complex set of interactions may determine the ultimate outcome of fumonisin exposure.

Curiously, we find Wnt and Hippo signaling to be the common threads throughout our data, even though it is ceramide synthase and sphingolipid biosynthesis which are directly inhibited by FB1. This trend spans both our mouse and human studies, suggesting that similar mechanisms mediate the effects of FB1 in both organisms. Furthermore, many of these genes are themselves NTD candidates, providing additional support for a role of FB1 in human NTD pathogenesis.

Chapter 5: Synthesis of findings, implications, and future directions

5.1 NTDs as complex disease

As evidenced by the range of phenotypes, only briefly described here (Section 1.2.1), and the panoply of contributing risk factors known and possibly still unknown (Sections 1.3, 1.4), NTDs "are not one disease but many" (Wallingford et al., 2013). Our findings reflect and support this conclusion: interactions of genetic factors (Section 2.1), environmental or epigenetic factors (Section 3.1), and interactions between genetic and environmental factors (Sections 2.2, 4.1, 4.2) may all contribute to NTD pathogenesis. Deciphering which mechanisms or factors are common to all NTDs, if any, and what is unique to the etiology of each phenotype or even each individual case is an ongoing challenge in NTD research.

To date, most human NTD studies have featured small sample sizes and a concomitant lack of statistical power (reviewed in Au et al., 2010). As discussed by Au et al., this is a natural consequence of the difficulty in ascertaining NTD cases, particularly for severe phenotypes. However, even when a single research group has collected enough samples for a sufficiently powered case-control genetic study, they have largely focused on rare variants and putative mutations, even though the multifactorial model also supports contribution of common variants to risk (reviewed in

Juriloff and Harris, 2012). Such approaches are likely to miss the complexity of interactions that underlie NTD etiology.

We propose pedigree-based high-throughput sequencing experiments as a means for identifying interactions that contribute specifically to human NTD etiology. In sporadic cases, each individual is likely to possess a unique set of risk variants; with over 240 candidate genes already implicated in NTD etiology (Harris and Juriloff, 2007, 2010), the potential genetic heterogeneity is considerable. However, affecteds in a multiplex pedigree are most likely to have a common pathogenesis, and through the evaluation of genetic sharing, a 'module' of interacting variants – such as we observed in Dvl-Rho complex components of one family (Section 2.1.2.4) – can be identified and functionally characterized. The variants and genes of these modules can then be queried in sufficiently powered case-control studies of sporadic cases to determine how frequently a given genetic module contributes to human NTD risk. The Hereditary Basis of Neural Tube Defects study at Duke has been ongoing since 1994, and currently includes 5,176 individuals in 1,366 families spanning all NTD phenotypes. Furthermore, our anencephaly subset includes 1,128 individuals (330 cases) from 351 families, about 20% of which are multiplex families that contain multiple affected individuals. Our cohort therefore provides an excellent basis for additional investigations of gene-gene interaction in human NTDs.

5.1.1 Functional characterization of variants in PCP genes

In our exome sequenced multiplex anencephaly family, we identified an aggregation of variants that we posit collectively attenuated PCP signaling in the affecteds below the level necessary for neural tube closure (Section 2.1.2.4). However, we did not have RNA available for this family, and could not interrogate gene transcript or isoform expression. To determine what functional impact these variants have, we propose introducing equivalent mutations *in vitro* using a CRISPR-Cas9 site-directed mutagenesis system and use real-time PCR to evaluate gene and isoform expression based upon our predictions of variant splicing effects. We are currently identifying an appropriate cell line to use, i.e. one with strong expression of PCP signaling genes. Subsequently, gene constructs for variants with validated functional effects will be introduced into a *Xenopus* model system in collaboration with the Wallingford lab at the University of Texas in Austin to evaluate impacts on neural tube closure, a group with past experience in PCP signaling and *Xenopus* models of NTDs (Park et al., 2006; Wallingford and Harland, 2002).

5.1.2 Wnt genes as modifiers of environmental risk

Interestingly, we found that SNPs in Wnt genes were also some of the most strongly associated with blood sphingolipid levels in humans exposed to fumonisin (Section 4.2.2.2). These findings highlight Wnt signaling genes as potential mediators of

environmental risk, particularly in the context of FB1 exposure, an aspect which to our knowledge has not yet been explored. The genetic variation we observed in Wnt genes of LM/Bc and SWV mice, such as *Lrp2*, *Lrp6, Celsr1,* and *Fzd1* (Section 2.2.2.5) may provide a ready context for further investigation of these gene-environment interactions. Namely, we can assay gene expression directly in control and treated embryos, and also introduce single or multiple variants from the LM/Bc strain into an *in vitro* system to evaluate impacts on an individual and collective basis. We will then be able to evaluate these variants for potential splicing effects and the relative expression of isoforms, as with the human PCP variants above.

5.1.3 TEAD1 and Hippo signaling

Our study of fumonisin as a potential NTD risk factor naturally began with a focus on sphingolipid metabolism (Gelineau-van Waes et al., 2005, 2012); however, components of the Hippo signaling pathway appeared repeatedly in our data, from analyses of mouse genetic variation (Section 2.2.2.5), human SNP associations (Section 4.2.2.2) and mouse epigenetic modifications (Section 4.1.2.4). We also identified a heterozygous novel variant in *TEAD1* among members of our multiplex anencephaly family (Section 2.1.2.5), and a possible site of differential methylation in discordant MZ twins (Section 3.1.2.6), although the differential methylation did not validate. These data suggest that *TEAD1* may be of particular interest in human NTD etiology.

TEAD family transcription factors are the major targets of Hippo signaling, which cross talks extensively with Shh, Wnt, BMP/TGF-β, and Notch signaling pathways (reviewed in Zhao et al., 2010), all processes intimately involved in neural development. *Tead1* is an established NTD candidate gene, identified from a *Tead1/2* double-knockout mouse strain (Sawada et al., 2008). Mice null for *Tead1* alone exhibit heart development defects and die around E11.5 (Chen et al., 1994), while *Tead2* single-knockouts exhibit no overt phenotype (Sawada et al., 2008), suggesting partial redundancy similar to that of Dvl family members (Etheridge et al., 2008). By extension, *TEAD1* and its interactors may comprise another important genetic module in NTD pathogenesis, one that to our knowledge has received little attention. We propose moving forward with the functional characterization of genetic variation in LM/Bc *Tead1* and its interactors through analysis of allele-specific expression analyses and *in vitro* mutagenesis within strain-derived cell lines. Human variants will be selected for functional analysis based upon findings in the model system.

5.2 Fumonisins as NTD risk factors

To date, little is known about fumonisin toxicity in humans, or its contribution to human NTD risk. However, evidence supports FB1 as a candidate risk factor, from correlations between *F. vertillicioides* occurrence and NTD incidence to a mouse model for FB1-induced NTDs (Section 1.4.2.1). Our team's data supports that fumonisin FB1

exposure in humans has similar consequences for sphingolipid metabolism as we observe in the mouse model (Torres et al., 2014). In future studies, we intend to determine whether other findings from the mouse model hold up in the human context as well.

5.2.1 FB1 response profiles and genetic variation in mice and humans

With the addition of AKR mice to our team's model system, we now have strains representing minimal, moderate, and strong susceptibility to FB1-induced NTDs. With metabolic and gene expression profiles of these strains, we can extrapolate characteristics distinctive to each phenotype group, both at baseline and upon exposure to fumonisin. By expanding on our existing human data, particularly with the addition of gene expression profiles, we can determine how human biologic responses to fumonisin compare to those observed in the mice. If human sphingolipid metabolism and gene expression profiles show similar trends, these data would provide a strong indication that FB1 affects humans much the same as it does mice, and that fumonisin exposure is a risk factor contributing to human NTD pathogenesis.

5.2.2 Differential methylation relating to FB1 exposure

The Gelineau-van Waes lab is presently developing strain-specific cell lines that will permit the use of *in vitro* techniques for functional validation. Luciferase reporter constructs with partial or complete enzymatic methylation have previously been used

for functional analysis of DMRs (Dayeh et al., 2014; Unoki and Nakamura, 2003). We propose to validate our mouse DMRs with similar constructs, using unmethylated, partially methylated, and fully methylated versions in control and FB1-treated cells of either strain. This will allow us to investigate 1) whether a given methylation state for a construct elicits similar outcomes in LM/Bc- and SWV-derived cells; 2) if any functional effects of differential methylation also depend upon FB1; and 3) whether differential methylation is affected by strain specific genetic sequence (i.e. by incorporating variants from the other strain into constructs where appropriate). Given our observation that methylation changes are associated with FB1 exposure *in vivo*, we may also find that FB1 treatment *in vitro* leads to changes of methylation in the constructs, a possibility which would have to be verified or ruled out before proceeding with functional analysis.

Ultimately, the goal will be to translate this work with differential methylation into the human context. Others have previously used the HumanMethylation450k array on DNA extracted from dried blood spots (Joo et al., 2013) , not unlike those we have already collected for our genetic studies of samples collected in Guatemala. If we can obtain methylation profiles from our existing samples, we can incorporate these data into FB1 response profiles with expression and metabolite data. With these samples, collected as whole blood, we would be able to profile with the SureSelect Human Methyl-seq XT platform, allowing more direct comparisons. To truly compare these to

the mouse profiles, however, we would also need to obtain methylation data for mature female mice with chronic dietary fumonisin intake. Such a study would also allow us to compare embryonic profiles with adult and determine what effects, if any, are specific to life stage.

One additional avenue that was not pursued by our preliminary research is that in our model, some 20% of FB1-treated LM/Bc mouse embryos do not develop exencephaly (Gelineau-van Waes et al., 2005). As inbred mice are known to vary stochastically in epigenetic profile (Li et al., 2011a), these may possess an epigenetic difference (i.e. an 'epimutaton') that protects against FB1-induced NTDs. Regardless, profiles from such mice will also provide an additional filter for distinguishing epigenetic modifications with potential functional relevance to neural tube closure.

5.3 Epigenetic modifications and NTD etiology

The role of the epigenome in NTD etiology remains a chicken-and-egg quandary: do epigenetic changes have functional effects that contribute to NTD causation, or are the observed differences merely a consequence arising from other, more central problems (e.g. folate deficiency, diabetes)? We have documented DNA methylation differences in humans and LM/Bc mice with NTDs (Sections 3.1, 4.1), and the Gelineau-van Waes lab has also observed differences in the abundance of histone modifications after treatment with fumonisin, particularly H2BK12ac, H3K9ac, H4K5ac, H3K4me3,

and H3K27me (data not yet published); however, none of these observations resolve the question of causality. A proposal for functional validation in the mouse model system has been described above (Section 5.2.2), but further characterization remains to be done in the human context.

5.3.1 Further study of discordant monozygotic twins

A difficulty with the use of discordant DZ twins for validation of epigenetic differences in MZ twins, as in this research (Section 3.1.2.6), is that these twin pairs differ on both epigenetic and genetic levels. Just as genetic factors may be heterogeneous between unrelated cases, it is reasonable to expect that epigenetic factors may be as well. Accordingly, in future epigenetic research, we propose a similar pedigree-based profiling and comparison approach as for genetic studies. In our ongoing ascertainment of NTD cases, our study has collected two additional monozygotic twin pairs discordant for anencephaly. We were not able to include them in the research covered by this dissertation, but they will provide a basis for future investigations.

In such an experiment, we would generate methylome profiles using a high-throughput sequencing platform such as the Sure-Select Human Methyl-Seq XT kit (Agilent Technologies). This kit covers 85 Mb of the human genome, targeting regulatory regions as described for the mouse version (Section 4.1.1.2), and provides quantitative, single-base resolution of methylation levels. Profiles for unaffected

relatives as well as the unaffected twin will give a basis for filtering out naturally variable sites and changes less likely to contribute to NTD pathogenesis. For validation of site functionality, rather than looking to other human NTD cases with their likely different etiologies, we would employ *in vitro* assays described above. If we identify epigenetic modifications shared by several cases, either at individual CpG or gene level, then those would be candidates for screening in larger case-control cohorts as potentially more common elements of NTD etiology.

5.3.2 Functional role of mir-886

Another intriguing finding among our epigenetic studies is the hypomethylation of *mir-886* in some NTD cases (Section 3.1.2.7). The long form of this non-coding RNA has been shown to bind protein kinase R (Jeon et al., 2012), while its derivatives *mir-886-3p* and *mir-886-5p* may have regulatory functionality similar to other miRNAs. This gene is not present in mouse, *Xenopus*, or zebrafish, which pose both difficulties and opportunities for validation of its functionality in neural tube closure. Introducing the full-length and processed forms of this ncRNA into zebrafish or *Xenopus* systems would provide a ready test of whether excess *mir-886* promotes NTDs, without confounding by any endogenous version of the gene.

However, targets of these miRNAs may not be conserved, especially in species that do not possess an ortholog of *mir-886* to begin with. Determination of targets will

be best done in human cells with tandem affinity purification (Nonne et al., 2010). In this approach, cells stably expressing Flag-tagged Argonaute proteins would be transfected with biotinylated *mir-886-3p* or *mir-886-5p*. Anti-Flag beads are used to purify Argonaute complexes from cell extracts, and streptavidin beads to pull down the miRNA of interest and its bound targets. The enriched RNA can then be deep-sequenced and analyzed to identify targets of *mir-886*. This two-step procedure produces lower background than direct pulldown of the biotinylated miRNA, and allows direct identification of mRNAs physically bound by the miRNAs. Further *in vitro* constructs can then be performed to characterize the effects of *mir-886* binding upon specific target RNAs.

Appendix A: Reference Gene Sets

A.1 Genes associated with abnormal neural tube morphology in mice

This gene list was defined by the MGI database phenotype entry "abnormal neural tube morphology/development" (MP:0002151) and with child terms encompasses a total of 527 genes. This phenotype is by definition "any structural anomaly of or development of the hollow epithelial tube found on the dorsal side of the vertebrate embryo that develops into the central nervous system (i.e. brain and spinal cord)." Child terms include phenotypes such as craniorachischisis and spina bifida, but do not include exencephaly. Gene associations retrieved through MGI 5.15 on 10/30/13.

A.2 Genes associated with exencephaly in mice

This gene list was defined by the MGI database phenotype entry "exencephaly" (MP:0000914) and encompasses a total of 282 genes. This phenotype is by definition "neurocranial defects resulting in exposure or extrusion of the brain." Although the MGI phenotype hierarchy does not list exencephaly as a child term of "abnormal neural tube morphology," 172 (61%) of these genes are also associated with that phenotype. Gene associations retrieved through MGI 5.15 on 10/30/13.

A.3 Hedgehog signaling pathway

This gene list was defined by the Gene Ontology Biological Process term "smoothened signaling pathway" (GO:0007224). The list used is inclusive of genes associated in either human or mouse, and inclusive of child terms encompasses a total of 121 genes. Gene associations retrieved through AmiGO 1.8 on 06/08/14.

A.4 Wnt signaling pathway

This gene list was defined by the Gene Ontology Biological Process term "Wnt signaling pathway" (GO:0016055), and includes all branches of Wnt signaling. The list used is inclusive of genes associated in either human or mouse, and with child terms encompasses a total of 361 genes. Gene associations retrieved through AmiGO 1.8 on 06/08/14.

A.4.1 Wnt-PCP signaling pathway

This gene list was defined from literature searches, reviews on PCP signaling, and personal communication with Dr. John Wallingford. The included genes are *ANKRD6*, *CELSR1*, *CELSR2*, *CELSR3*, *DAAM1*, *DVL1*, *DVL2*, *DVL3*, *FUZ*, *FZ1*, *FZ2*, *FZ3*, *FZ6*, *HECW1*, *INTU*, *LRP6*, *NOTCH1*, *PRICKLE1*, *PRICKLE2*, *PRICKLE3*, *PRICKLE4*, *PTK2*, *RSG1*, *VANGL1*, *VANGL2*, and *WDPCP*.

A.5 Sphingolipid metabolism genes

This gene list was defined by the Gene Ontology Biological Process term "sphingolipid metabolic process" (GO:0006665). The list used is inclusive of genes associated in either human or mouse, and with child terms encompasses a total of 135 genes. GDF1 is not included. Gene associations retrieved through AmiGO 1.8 on 06/08/14.

Appendix B: Primer Sequences

B.1 Human PAX6 and TEAD1 validation primers

These primers were used to validate novel variants as described in Section 2.1.1.6. Products were amplified using a touchdown protocol of 64-58 C.

Table 26. Primer sequences for validation of novel variants in *PAX6* and *TEAD1*.

Region	Forward Primer	Reverse Primer
PAX6 intron 6	AAGTGCTGGACAATCAAAACG	GGTGGGAGGAGGTAAAGAGG
TEAD1 intron 10	GAACCATTTCCCCTCCAGA	TAAAAAGCCCCAGCATCATC

B.2 Mouse Cers1-Gdf1 exon 8 validation primers

These primers were used to validate rs32650885 and rs47220823 in representative LM/Bc and SWV mice as described in Section 2.2.1.2. Products were amplified using a touchdown protocol of 64-58 C.

Table 27. Primer sequences for validation of LM/Bc and SWV variants in *Cers1-Gdf1* exon 8.

Function	Forward Primer	Reverse Primer
PCR amplification	CTCCCTGCTGCTGGTGAC	TCTCGTGATTGAGGGGCTAC
Nested sequencing primer		ACCTGAGAGATGCGAGAAGC

B.3 Human methyl validation primers

B.3.1 Primer sequences for the EpiTyper assay

These primers were used to amplify target DMRs from the discordant MZ twin pair for validation by EpiTyper assay. Products were amplified following the kit manufacturer's protocol.

Table 28. Primers used in the EpiTyper assay for validation of human DMRs. EpiTyper-specific tags given in lowercase.

Gene	Forward	Reverse
APC	aggaagagagTGTGAATATTGATTATGTTGTTTAGTAGT	cagtaatacgactcactatagggagaaggctTCACATCAAAAATCCAAATTAAATC
RAP2B	aggaagagagTGTAGTTTTGTATAAAGAAAATAGGAGG	cagtaatacgactcactatagggagaaggctATACAAAAACTTCCTACTTCCAAACC
DAAM2	aggaagagagGTGGAGTTAAGGATATTTAGGTTTG	cagtaatacgactcactatagggagaaggctTTTAATCCTAACCTCCCAAAAAAAC
CAMTA	aggaagagagTTGGGTTTTATTTAATATTGTGATTTG	cagtaatacgactcactatagggagaaggctACTATCTAAAAACCTCCCTTAATCC
GLI3	aggaagagagGGAGATTAATAGGATATGTGGGTTTT	cagtaatacgactcactatagggagaaggctAACTCTTAAAAATTCAAAAACAAACTCAA
miR886	aggaagagagAGTTTGGTTGTTGGATTTAGGTAGA	cagtaatacgactcactatagggagaaggctACCCCAACACAAAAATAAACAAATA
LRP1B	aggaagagagTGAGGTATTTGGTGAGTGTAGAGTATT	cagtaatacgactcactatagggagaaggctCCCTAACAATAAACACATAAACCTATTT

B.3.2 Primer sequences for targeted bisulfite resequencing

These primers were used to amplify target DMRs from the discordant MZ twin pair for validation by high-throughput sequencing (Section 3.1.1.5). All pairs were optimized on a standard PCR program (not touchdown), with annealing temperatures and Mg^{2+} concentrations specific to each amplicon.

Table 29. Primers used in targeted bisulfite resequencing for validation of human DMRs.

Region	Forward	Reverse	Mg^{2+}	Anneal
APC2	AGTTTAGTTGTATGTAAGTGGGTGT	ACTAACCTACCCATCCCTTC	3.5 mM	58 C
APC2	GTTGATGGATAGTGATTGTTATTG	TACACCAAACAAATAACTTTACATT	3.5 mM	54 C
BCOR	ATAGTTGGGTTTTATGTAAGGTTTA	CCACCACCACCACTACTACTAA	1.5 mM	55 C
GRM8	TTTTGAAATAGTAAGTAATGGAGGA	TCACACACTAACAATCTACACAAAC	3.5 mM	56 C
GRM8	GATTGTGTTTGGGATTGTTAAT	CACAACCATATTCCTTCTTCTAC	3.5 mM	54 C
HES7	TAGTTGAGATAGGGAGTAATTTGAG	TAACTACTTAATATCACATAATCTTTTCC	3.5 mM	51 C
KIRREL3	AATAAGGTAGGAAAGATTTGTGAG	ACTAAAACTAACTACCAACCCTACA	3.5 mM	58 C
MED12	TTATTTTGTGTGGTTTTTATTATGT	ACAAATCATCCTAAATAACTACCAA	3.5 mM	53 C
MSX2	TTGTTTGAAGTTATAGAATGGGTAG	ACACAATAAACAATCACAAACTATC	3.5 mM	51 C
NCOR2	TGTGAGAAAGTTTAGGTAGGATG	CTCCCTCTCTATCAAATAAATAACA	3.5 mM	55 C
NR1H3	TAAGAGTTGTTAAGAGTGTTGGGTA	TCCTTCCTACCTAATAACCACA	1.5 mM	55 C
RARA	TTTGATGTTGAGTAGTATTAGGG	ACACAACACAAACAAACTCAC	3.5 mM	57 C
SMPD3	ATTTAGTATTTGGAGTGGTGAATAG	TTACCTACCCACAACATAAACC	3.5 mM	56 C
SPTBN4	TTTTTAGAAGGATTGAGTTTTGT	ACATTCTCCAAATACACACACTAC	2.5 mM	52 C
TEAD1	TGGTGTTTAGATTTTGTTGTTG	AAACAATACAACATCATCTTCC	3.5 mM	53 C

ZBTB16	TTAGTGGTTGGTTTTAGAGGAG	CCTAAACTTTCTCATATTAAATTCC	3.5 mM	56 C
miR886	GTGGGTTTTGAATTTTAGTATAGAG	AAAACACATTAACAAAACACCTAA	3.5 mM	54 C
miR886	AATTATAGAAGAGTGATGATGTGG	AAATACTCCTCAACAACCTAACTAC	3.5 mM	55 C
miR886	TTTTATTTGGTAGTATAGGTTGGTT	CATACCAAACTTTCTATCTATCCAT	3.5 mM	54 C
miR886	TTATTTGTTGTGGTTTAATAGGTAA	AACCACTAATCTTACAAAACATACA	3.5 mM	54 C

B.4 Mouse methyl validation primers

These primers were used to amplify target DMRs identified in an FB1-treated LM/Bc mouse embryo for validation by high-throughput sequencing (Section 4.1.1.4). All pairs were optimized on a standard PCR program (not touchdown), with annealing temperatures and Mg^{2+} concentrations specific to each amplicon.

Amplicon	Forward	Reverse	Mg^{2+}	Anneal
CTNND2_A	TTAGAGTAAAGTTTTTAGGAAGGAG	ATACAACCTTACAAAATTAAAACAC	2.5 mM	54 C
CTNND2_B	ATTTAGTGGAAGATTGTTTGATATG	CCTTCTATATTCCAACAACTAATTC	2.5 mM	51 C
E2F2	TAGGTTTTTGGAAGTAATGAGTAGA	ATCTATTCCCCAAACACAATC	3.5 mM	55 C
FAT3	TTAAGTTTAGGAAGAAGATGAAGG	CCACAAACTATACTTCTAACTTTCC	3.5 mM	55 C
FRMD4A	TTAAGAATAAATTGTTGTGTGTAGG	TACACAAACCTCAACCCTCTC	3.5 mM	56 C
L3MBTL1_A	GATATTAAAAAGTGGTGGGTTT	ATACCTTCTACACTTCCATTAAACA	2.5 mM	50 C
L3MBTL1_B	GTTATTGAATATGATAGGAGTGGAG	AAAAACCACTAAACCTCAATCTT	3.5 mM	55 C
L3MBTL1_C	GAGTTTGAGTGTGTTGTTTATGG	TAATCCACATTCTCCTAAAACTAAC	3.5 mM	56 C
NF2	TGTATGAGGTATTGTTGTAGTTAGTG	ATTAAACAAATTTCAATTCCTCAAC	2.5 mM	55 C

PRICKLE1_A	ATTTTGAAAAAGAATAAAGAAGTGG	ATATAATCCACCTCAACAAATAACA	3.5 mM	54 C
PRICKLE1_B	GTTATTTTTATGGAGTTGTTTTGA	CAACTAACCTTTTTCTAACTAAACC	3.5 mM	55 C
SHMT2_A	TTTTTGATTGATTGTTAGGTGTT	ACTAATAAATAACTTCAAAAATCCT	2.5 mM	50 C
SHMT2_B	TTGTTAGAGTTTATAGGTTGTAGGG	TTCCCTCTTCTCTATACTATCCTCT	3.5 mM	53 C
SMAD6	TAGTAGGGAAAGAAGTTATGGTTT	ATAAATCCTCCTACCTCAAAATACA	3.5 mM	56 C
TEAD4	GGTGAAATAAATTTGGGTTAGTAG	ATACATCCAAATAAAAACTACATCC	3.5 mM	52 C
VGLL3	TGGTTTATTTTTAATGGAGGAA	ATTAAAACAAAATTTACAACCCTCT	3.5 mM	53 C
VGLL4_A	GTGGTGATTGTTGAGTGTAGAG	TTAACAATAACTACATACCCAAACC	2.5 mM	58 C
VGLL4_B	ATGTAATGATTTTGTTAGATTTTGT	CTCTTTACTTTTATAATAACCACTCC	2.5 mM	55 C
WWC2_A	ATGTTGGATATTGAGTTGTAGTTGT	CAAAACATACTACCTTCCTCTTTAC	3.5 mM	56 C
WWC2_B	GGGATAAAGAGGATTTATAGTATGTT	ACTTCACTTTAACACTTCACTCAAC	3.5 mM	52 C

References

Agopian, A.J., Tinker, S.C., Lupo, P.J., Canfield, M.A., Mitchell,, L.E., and Study, and the N.B.D.P. (2013). Proportion of neural tube defects attributable to known risk factors. Birt. Defects Res. A. Clin. Mol. Teratol. *97*, 42–46.

Alvarez, I.S., and Schoenwolf, G.C. (1992). Expansion of surface epithelium provides the major extrinsic force for bending of the neural plate. J. Exp. Zool. *261*, 340–348.

Alvarez-Medina, R., Cayuso, J., Okubo, T., Takada, S., and Martí, E. (2008). Wnt canonical pathway restricts graded Shh/Gli patterning activity through the regulation of Gli3 expression. Development *135*, 237–247.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Au, K.S., Ashley-Koch, A., and Northrup, H. (2010). Epidemiologic and genetic aspects of spina bifida and other neural tube defects. Dev. Disabil. Res. Rev. *16*, 6–15.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In Current Protocols in Bioinformatics, (John Wiley & Sons, Inc.), pp. 11.10.1–11.10.33.

Beaudin, A.E., and Stover, P.J. (2009). Insights into Metabolic Mechanisms Underlying Folate-Responsive Neural Tube Defects: A Mini-Review. Birt. Defects Res. A. Clin. Mol. Teratol. *85*, 274–284.

Bengtsson, H., Epifantseva, I., Abrink, M., Kylberg, A., Kullander, K., Ebendal, T., and Usoskin, D. (2008). Generation and Characterization of a Gdf1 Conditional Null Allele. Genesis *46*, 368–372.

Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., and Gunderson, K.L. (2009). Genome-wide DNA Methylation Profiling Using Infinium Assay. Epigenomics *1*, 177–200.

Bilke, S., Triche, T., and Bootwalla, M. (2012). methylumi: Handle Illumina methylation data.

Bonner, J., Gribble, S.L., Veien, E.S., Nikolaus, O.B., Weidinger, G., and Dorsky, R.I. (2008). Proliferation and patterning are mediated independently in the dorsal spinal cord downstream of canonical Wnt signaling. Dev. Biol. *313*, 398–407.

Burren, K.A., Scott, J.M., Copp, A.J., and Greene, N.D.E. (2010). The Genetic Background of the Curly Tail Strain Confers Susceptibility to Folate-Deficiency-Induced Exencephaly. Birt. Defects Res. A. Clin. Mol. Teratol. *88*, 76–83.

Byun, H.-M., Siegmund, K.D., Pan, F., Weisenberger, D.J., Kanel, G., Laird, P.W., and Yang, A.S. (2009). Epigenetic Profiling of Somatic Tissues from Human Autopsy Specimens Identifies Tissue- and Individual-specific DNA Methylation Patterns. Hum. Mol. Genet. *18*, 4808–4817.

Cafici, D., and Sepulveda, W. (2003). First-Trimester Echogenic Amniotic Fluid in the Acrania-Anencephaly Sequence. J. Ultrasound Med. *22*, 1075–1079.

Canfield, M.A., Collins, J.S., Botto, L.D., Williams, L.J., Mai, C.T., Kirby, R.S., Pearson, K., Devine, O., Mulinare, J., and for the National Birth Defects Prevention Network (2005). Changes in the birth prevalence of selected birth defects after grain fortification wiht folic acid in the United States: Findings from a multi-state population-based study. Birt. Defects Res. A. Clin. Mol. Teratol. *73*, 679–689.

Carlson, C.S., Eberle, M.A., Rieder, M.J., Yi, Q., Kruglyak, L., and Nickerson, D.A. (2004). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am. J. Hum. Genet. *74*, 106–120.

Cattanach, B.M., Rasberry, C., Evans, E.P., and Woodward, A.M. (1996). Two Sey (Pax6) mutants. Mouse Genome *94*, 978.

Cengiz, B., Söylemez, F., Öztürk, E., and Çavdar, A.O. (2004). Serum zinc, selenium, copper, and lead levels in women with second-trimester induced abortion resulting from neural tube defects. Biol. Trace Elem. Res. *97*, 225–235.

Chang, H., Zhang, T., Zhang, Z., Bao, R., Fu, C., Wang, Z., Bao, Y., Li, Y., Wu, L., Zheng, X., et al. (2011). Tissue-specific distribution of aberrant DNA methylation associated with maternal low-folate status in human neural tube defects. J. Nutr. Biochem. *22*, 1172–1177.

Chen, X., Guo, J., Lei, Y., Zou, J., Lu, X., Bao, Y., Wu, L., Wu, J., Zheng, X., Shen, Y., et al. (2010). Global DNA hypomethylation is associated with NTD-affected pregnancy: a case-control study. Birth Defects Res. Part A *88*, 575–581.

Chen, Y.-A., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., and Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics Off. J. DNA Methylation Soc. *8*.

Chen, Z., Friedrich, G.A., and Soriano, P. (1994). Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. Genes Dev. *8*, 2293–2301.

Chesnutt, C., Burrus, L.W., Brown, A.M.C., and Niswander, L. (2004). Coordinate regulation of neural tube patterning and proliferation by TGFβ and WNT activity. Dev. Biol. *274*, 334–347.

Cockroft, D.L., Brook, F.A., and Copp, A.J. (1992). Inositol deficiency increases the susceptibility to neural tube defects of genetically predisposed (curly tail) mouse embryos in vitro. Teratology *45*, 223–232.

Copp, A.J., Brook, F.A., and Roberts, H.J. (1988). A cell-type-specific abnormality of cell proliferation in mutant (curly tail) mouse embryos developing spinal neural tube defects. Development *104*, 285–295.

Davies, M.N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Coarfa, C., Harris, R.A., Milosavljevic, A., Troakes, C., et al. (2012). Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol. *13*, R43.

Dayeh, T., Volkov, P., Salö, S., Hall, E., Nilsson, E., Olsson, A.H., Kirkpatrick, C.L., Wollheim, C.B., Eliasson, L., Rönn, T., et al. (2014). Genome-Wide DNA Methylation Analysis of Human Pancreatic Islets from Type 2 Diabetic and Non-Diabetic Donors Identifies Candidate Genes That Influence Insulin Secretion. PLoS Genet *10*, e1004160.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. *37*, e67.

Eikel, D., Lampen, A., and Nau, H. (2006). Teratogenic Effects Mediated by Inhibition of Histone Deacetylases: Evidence from Quantitative Structure Activity Relationships of 20 Valproic Acid Derivatives. Chem. Res. Toxicol. *19*, 272–278.

Elwood, J.M., Little, J., and Elwood, J.H. (1992). Epidemiology and Control of Neural Tube Defects (Oxford: Oxford University Press).

ENCODE Project Consortium (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. *9*, e1001046.

Etheridge, S.L., Ray, S., Li, S., Hamblet, N.S., Lijam, N., Tsang, M., Greer, J., Kardos, N., Wang, J., Sussman, D.J., et al. (2008). Murine Dishevelled 3 Functions in Redundant Pathways with Dishevelled 1 and 2 in Normal Cardiac Outflow Tract, Cochlea, and Neural Tube Development. PLoS Genet *4*, e1000259.

Ezin, A.M., Fraser, S.E., and Bronner-Fraser, M. (2009). Fate Map and Morphogenesis of Presumptive Neural Crest and Dorsal Neural Tube. Dev. Biol. *330*, 221–236.

Finnell, R.H., Moon, S.P., Abbott, L.C., Golden, J.A., and Chernoff, G.F. (1986). Strain differences in heat-induced neural tube defects in mice. Teratology *33*, 247–252.

Finnell, R.H., Bennett, G.D., Karras, S.B., and Mohl, V.K. (1988). Common hierarchies of susceptibility to the induction of neural tube defects in mouse embryos by valproic acid and its 4-propyl-4-pentenoic acid metabolite. Teratology *38*, 313–320.

Fleming, A., and Copp, A.J. (1998). Embryonic folate metabolism and mouse neural tube defects. Science *280*, 2107–2109.

Fleming, A., and Copp, A.J. (2000). A genetic risk factor for mouse neural tube defects: defining the embryonic basis. Hum. Mol. Genet. *9*, 575 –581.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl. Acad. Sci. U. S. A. *102*, 10604–10609.

Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Mark J. Rieder, Altshuler, D., Shendure, J., et al. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216–220.

Gelineau-van Waes, J., and Finnell, R.H. (2001). Genetics of neural tube defects. Semin. Pediatr. Neurol. *8*, 160–164.

Gelineau-van Waes, J., Starr, L., Maddox, J., Aleman, F., Voss, K.A., Wilberding, J., and Riley, R.T. (2005). Maternal fumonisin exposure and risk for neural tube defects: mechanisms in an in vivo mouse model. Birt. Defects Res. A. Clin. Mol. Teratol. *73*, 487–497.

Gelineau-van Waes, J., Rainey, M.A., Maddox, J.R., Voss, K.A., Sachs, A.J., Gardner, N.M., Wilberding, J.D., and Riley, R.T. (2012). Increased sphingoid base-1-phosphates and failure of neural tube closure after exposure to fumonisin or FTY720. Birt. Defects Res. A. Clin. Mol. Teratol. *94*, 790–803.

Gilbert, S.F. (2000). Developmental Biology (Sinauer Associates).

Gray, J.D., Kholmanskikh, S., Castaldo, B.S., Hansler, A., Chung, H., Klotz, B., Singh, S., Brown, A.M.C., and Ross, M.E. (2013). Lrp6 Exerts Non-Canonical Effects on Wnt Signaling During Neural Tube Closure. Hum. Mol. Genet. ddt277.

Greene, N.D.E., and Copp, A.J. (2009). Development of the Vertebrate Central Nervous System: Formation of the Neural Tube. Prenat. Diagn. *29*, 303–311.

Gregory, S.G., Connelly, J.J., Towers, A.J., Johnson, J., Biscocho, D., Markunas, C.A., Lintas, C., Abramson, R.K., Wright, H.H., Ellis, P., et al. (2009). Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. BMC Med. *7*.

Grubb, S.C., Bult, C.J., and Bogue, M.A. (2014). Mouse Phenome Database. Nucleic Acids Res. *42*, D825–D834.

Habas, R., Kato, Y., and He, X. (2001). Wnt/Frizzled activation of Rho regulates vertebrate gastrulation and requires a novel Formin homology protein Daam1. Cell *107*, 843–854.

Hait, N.C., Allegood, J., Maceyka, M., Strub, G.M., Harikumar, K.B., Singh, S.K., Luo, C., Marmorstein, R., Kordula, T., Milstien, S., et al. (2009). Regulation of Histone Acetylation in the Nucleus by Sphingosine-1-Phosphate. Science *325*, 1254–1257.

Hamblet, N.S., Lijam, N., Ruiz-Lozano, P., Wang, J., Yang, Y., Luo, Z., Mei, L., Chien, K.R., Sussman, D.J., and Wynshaw-Boris, A. (2002). Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure. Dev. Camb. Engl. *129*, 5827–5838.

Harris, M.J., and Juriloff, D.M. (2007). Mouse Mutants With Neural Tube Closure Defects and Their Role in Understanding Human Neural Tube Defects. Birth Defects Res. Part A *79*, 187–210.

Harris, M.J., and Juriloff, D.M. (2010). An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. Birth Defects Res. Part A *88*, 653–669.

Harris, M.J., Juriloff, D.M., and Biddle, F.G. (1984). Cortisone cure of the lidgap defect in fetal mice: a dose-response and time-response study. Teratology *29*, 287–295.

Hebsgaard, S.M., Korning, P.G., Tolstrup, N., Engelbrecht, J., Rouzé, P., and Brunak, S. (1996). Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. Nucleic Acids Res. *24*, 3439–3452.

Heid, M.K., Bills, N.D., Hinrichs, S.H., and Clifford, A.J. (1992). Folate deficiency alone does not produce neural tube defects in mice. J. Nutr. *122*, 888–894.

Hendricks, K.A., Nuno, O.M., Suarez, L., and Larsen, R. (2001). Effects of hyperinsulinemia and obesity on risk of neural tube defects among Mexican Americans. Epidemiol. Camb. Mass *12*, 630–635.

Houdayer, C., Dehainault, C., Mattler, C., Michaux, D., Caux-Moncoutier, V., Pagès-Berhouet, S., d' Enghien, C.D., Laugé, A., Castera, L., Gauthier-Villars, M., et al. (2008). Evaluation of in silico splice tools for decision-making in molecular diagnosis. Hum. Mutat. *29*, 975–982.

Hu, N., Strobl-Mazzulla, P., Sauka-Spengler, T., and Bronner, M.E. (2012). DNA methyltransferase3A as a molecular switch mediating the neural tube-to-neural crest fate transition. Genes Dev. *26*, 2380–2385.

Huang, J., Wu, J., Li, T., Song, X., Zhang, B., Zhang, P., and Zheng, X. (2011). Effect of Exposure to Trace Elements in the Soil on the Prevalence of Neural Tube Defects in a High-Risk Area of China. Biomed. Environ. Sci. *24*, 94–101.

Huangfu, D., Liu, A., Rakeman, A.S., Murcia, N.S., Niswander, L., and Anderson, K.V. (2003). Hedgehog signalling in the mouse requires intraflagellar transport proteins. Nature *426*, 83–87.

Jeon, S.H., Lee, K., Lee, K.S., Kunkeaw, N., Johnson, B.H., Holthauzen, L.M.F., Gong, B., Leelayuwat, C., and Lee, Y.S. (2012). Characterization of the direct physical interaction of nc886, a cellular non-coding RNA, and PKR. FEBS Lett. *586*, 3477–3484.

Jin, L., Zhang, L., Li, Z., Liu, J., Ye, R., and Ren, A. (2013). Placental concentrations of mercury, lead, cadmium, and arsenic and the risk of neural tube defects in a Chinese population. Reprod. Toxicol. *35*, 25–31.

Joo, J.E., Wong, E.M., Baglietto, L., Jung, C.-H., Tsimiklis, H., Park, D.J., Wong, N.C., English, D.R., Hopper, J.L., Severi, G., et al. (2013). The use of DNA from archival dried blood spots with the Infinium HumanMethylation450 array. BMC Biotechnol. *13*, 23.

Juriloff, D.M., and Harris, M.J. (2012). A consideration of the evidence that genetic defects in planar cell polarity contribute to the etiology of human neural tube defects. Birt. Defects Res. A. Clin. Mol. Teratol. *94*, 824–840.

Kardon, T., Stroobant, V., Veiga-da-Cunha, M., and Schaftingen, E.V. (2008). Characterization of mammalian sedoheptulokinase and mechanism of formation of erythritol in sedoheptulokinase deficiency. FEBS Lett. *582*, 3330–3334.

Kingsley, D.M. (1994). The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes Dev. *8*, 133–146.

Kosoy, R., Nassir, R., Tian, C., White, P.A., Butler, L.M., Silva, G., Kittles, R., Alarcon-Riquelme, M.E., Gregersen, P.K., Belmont, J.W., et al. (2009). Ancestry Informative Marker Sets for Determining Continental Origin and Admixture Proportions in Common Populations in America. Hum. Mutat. *30*, 69–78.

Kozbial, P.Z., and Mushegian, A.R. (2005). Natural history of S-adenosylmethionine-binding proteins. BMC Struct. Biol. *5*, 19.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinforma. Oxf. Engl. *27*, 1571–1572.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lee, K., Kunkeaw, N., Jeon, S.H., Lee, I., Johnson, B.H., Kang, G.-Y., Bang, J.Y., Park, H.S., Leelayuwat, C., and Lee, Y.S. (2011). Precursor miR-886, a novel noncoding RNA repressed in cancer, associates with PKR and modulates its activity. RNA *17*, 1076–1089.

Lemos, M.C., Harding, B., Reed, A.A.C., Jeyabalan, J., Walls, G.V., Bowl, M.R., Sharpe, J., Wedden, S., Moss, J.E., Ross, A., et al. (2009). Genetic background influences embryonic lethality and the occurrence of neural tube defects in Men1 null mice: relevance to genetic modifiers. J. Endocrinol. *203*, 133–142.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754 –1760.

Li, L.-C., and Dahiya, R. (2002). MethPrimer: Designing Primers for Methylation PCRs. Bioinformatics *18*, 1427–1431.

Li, C.C.Y., Cropley, J.E., Cowley, M.J., Preiss, T., Martin, D.I.K., and Suter, C.M. (2011a). A Sustained Dietary Change Increases Epigenetic Variation in Isogenic Mice. PLoS Genet *7*, e1001380.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. *25*, 2078–2079.

Li, J.-H., Xiao, X., Zhang, Y.-N., Wang, Y.-M., Feng, L.-M., Wu, Y.-M., and Zhang, Y.-X. (2011b). MicroRNA miR-886-5p inhibits apoptosis by down-regulating Bax expression in human cervical carcinoma cells. Gynecol. Oncol. *120*, 145–151.

Li, N., Ye, M., Li, Y., Yan, Z., Butcher, L.M., Sun, J., Han, X., Chen, Q., zhang, X., and Wang, J. (2010). Whole genome DNA methylation analysis based on high throughput sequencing technology. Methods *52*, 203–212.

Li, Z., Ren, A., Zhang, L., Ye, R., Li, S., Zheng, J., Hong, S., Wang, T., and Li, Z. (2006). Extremely high prevalence of neural tube defects in a 4-county area in Shanxi Province, China. Birt. Defects Res. A. Clin. Mol. Teratol. *76*, 237–240.

Lie, R.T. (2006). An International Perspective on Anencephaly and Spina Bifida. In Neural Tube Defects: From Origin to Treatment, (Oxford: Oxford University Press), pp. 117–132.

Lim, K.H., and Fairbrother, W.G. (2012). Spliceman--a computational web server that predicts sequence variations in pre-mRNA splicing. Bioinforma. Oxf. Engl. *28*, 1031–1032.

Liu, H.X., Cartegni, L., Zhang, M.Q., and Krainer, A.R. (2001). A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat. Genet. *27*, 55–58.

Liu, X., Jian, X., and Boerwinkle, E. (2013). dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum. Mutat. *34*, E2393–2402.

Liu, Y., Balaraman, Y., Wang, G., Nephew, K.P., and Zhou, F.C. (2009). Alcohol exposure alters DNA methylation profiles in mouse embryos at early neurulation. Epigenetics *4*, 500–511.

Liu, Y., Siegmund, K.D., Laird, P.W., and Berman, B.P. (2012). Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data. Genome Biol. *13*, R61.

Maceyka, M., Harikumar, K.B., Milstien, S., and Spiegel, S. (2012). Sphingosine-1-phosphate signaling and its role in disease. Trends Cell Biol. *22*, 50–60.

Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963.

Marasas, W.F.O., Riley, R.T., Hendricks, K.A., Stevens, V.L., Sadler, T.W., Gelineau-van Waes, J., Missmer, S.A., Cabrera, J., Torres, O.A., Gelderblom, W.C.A., et al. (2004). Fumonisins disrupt sphingolipid metabolism, folate transport, and neural tube development in embryo culture and in vivo: a potential risk factor for human neural tube defects among populations consuming fumonisin-contaminated maize. J. Nutr. *134*, 711–716.

Marean, A., Graf, A., Zhang, Y., and Niswander, L. (2011). Folic acid supplementation can adversely affect murine neural tube closure and embryonic survival. Hum. Mol. Genet. ddr289.

Mastick, G.S., Davis, N.M., Andrew, G.L., and Easter, S.S. (1997). Pax-6 functions in boundary formation and axon guidance in the embryonic mouse forebrain. Development *124*, 1985–1997.

Matsumoto, A., Hatta, T., Moriyama, K., and Otani, H. (2002). Sequential observations of exencephaly and subsequent morphological changes by mouse exo utero development system: analysis of the mechanism of transformation from exencephaly to anencephaly. Anat. Embryol. (Berl.) *205*, 7–18.

Megason, S.G., and McMahon, A.P. (2002). A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. Dev. Camb. Engl. *129*, 2087–2098.

Mencarelli, C., and Martinez–Martinez, P. (2013). Ceramide function in the brain: when a slight tilt is enough. Cell. Mol. Life Sci. *70*, 181–203.

Menegola, E., Di Renzo, F., Broccia, M.L., Prudenziati, M., Minucci, S., Massa, V., and Giavini, E. (2005). Inhibition of histone deacetylase activity on specific embryonic tissues as a new mechanism for teratogenicity. Birth Defects Res. B. Dev. Reprod. Toxicol. *74*, 392–398.

Missmer, S.A., Suarez, L., Felkner, M., Wang, E., Merrill, A.H., Jr, Rothman, K.J., and Hendricks, K.A. (2006). Exposure to fumonisins and the occurrence of neural tube defects along the Texas-Mexico border. Environ. Health Perspect. *114*, 237–241.

Moore, L.L., Singer, M.R., Bradlee, M.L., Rothman, K.J., and Milunsky, A. (2000). A prospective study of the risk of congenital defects associated with maternal obesity and diabetes mellitus. Epidemiol. Camb. Mass *11*, 689–694.

Mort, M., Sterne-Weiler, T., Li, B., Ball, E.V., Cooper, D.N., Radivojac, P., Sanford, J.R., and Mooney, S.D. (2014). MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. Genome Biol. *15*, R19.

Mosley, B.S., Cleves, M.A., Siega-Riz, A.M., Shaw, G.M., Canfield, M.A., Waller, D.K., Werler, M.M., and Hobbs, C.A. (2009). Neural Tube Defects and Maternal Folate Intake Among Pregnancies Conceived After Folic Acid Fortification in the United States. Am. J. Epidemiol. *169*, 9–17.

Nonne, N., Ameyar-Zazoua, M., Souidi, M., and Harel-Bellan, A. (2010). Tandem affinity purification of miRNA target mRNAs (TAP-Tar). Nucleic Acids Res. *38*, e20–e20.

Nusse, R. (2012). Wnt Signaling. Cold Spring Harb. Perspect. Biol. *4*, a011163.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell *99*, 247–257.

Ollikainen, M., Smith, K.R., Joo, E.J.-H., Ng, H.K., Andronikos, R., Novakovic, B., Abdul Aziz, N.K., Carlin, J.B., Morley, R., Saffery, R., et al. (2010). DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal genome. Hum. Mol. Genet. *19*, 4176–4188.

Ou, C.Y., Stevenson, R.E., Brown, V.K., Schwartz, C.E., Allen, W.P., Khoury, M.J., Rozen, R., Oakley, G.P., and Adams, M.J. (1996). 5,10 Methylenetetrahydrofolate reductase genetic polymorphism as a risk factor for neural tube defects. Am. J. Med. Genet. *63*, 610–614.

Park, T.J., Haigo, S.L., and Wallingford, J.B. (2006). Ciliogenesis defects in embryos lacking inturned or fuzzy function are associated with failure of planar cell polarity and Hedgehog signaling. Nat. Genet. *38*, 303–311.

Parr, B.A., Shea, M.J., Vassileva, G., and McMahon, A.P. (1993). Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. Development *119*, 247–261.

Pederson, L., Ruan, M., Westendorf, J.J., Khosla, S., and Oursler, M.J. (2008). Regulation of bone formation by osteoclasts involves Wnt/BMP signaling and the chemokine sphingosine-1-phosphate. Proc. Natl. Acad. Sci. *105*, 20764–20769.

Pewzner-Jung, Y., Ben-Dor, S., and Futerman, A.H. (2006). When Do Lasses (Longevity assurance Genes) Become CerS (Ceramide Synthases)? J. Biol. Chem. *281*, 25001–25005.

Phillips, H.M., Murdoch, J.N., Chaudhry, B., Copp, A.J., and Henderson, D.J. (2005). Vangl2 Acts via RhoA Signaling to Regulate Polarized Cell Movements During Development of the Proximal Outflow Tract. Circ. Res. *96*, 292–299.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. Genetics *155*, 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. *81*, 559–575.

Rakyan, V.K., Beyan, H., Down, T.A., Hawa, M.I., Maslau, S., Aden, D., Daunay, A., Busato, F., Mein, C.A., Manfras, B., et al. (2011). Identification of Type 1 Diabetes–Associated DNA Methylation Variable Positions That Precede Disease Diagnosis. PLoS Genet *7*, e1002300.

Ray, J.G., Wyatt, P.R., Thompson, M.D., Vermeulen, M.J., Meier, C., Wong, P.-Y., Farrell, S.A., and Cole, D.E.C. (2007). Vitamin B12 and the risk of neural tube defects in a folic-acid-fortified population. Epidemiol. Camb. Mass *18*, 362–366.

Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. J. Comput. Biol. J. Comput. Mol. Cell Biol. *4*, 311–323.

Reimand, J., Arak, T., and Vilo, J. (2011). g:Profiler--a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res. *39*, W307–315.

Roelink, H., Porter, J.A., Chiang, C., Tanabe, Y., Chang, D.T., Beachy, P.A., and Jessell, T.M. (1995). Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. Cell *81*, 445–455.

Romanelli, V., Nakabayashi, K., Vizoso, M., Moran, S., Iglesias-Platas, I., Sugahara, N., Sugahara, N., Simón, C., Simón, C., Hata, K., et al. (2014). Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. Epigenetics *9*, 783–790.

Sawada, A., Kiyonari, H., Ukita, K., Nishioka, N., Imuta, Y., and Sasaki, H. (2008). Redundant Roles of Tead1 and Tead2 in Notochord Development and the Regulation of Cell Proliferation and Survival. Mol. Cell. Biol. *28*, 3177–3189.

Schatzmayr, G., and Streit, E. (2013). Global occurrence of mycotoxins in the food and feed chain: facts and figures. World Mycotoxin J. *6*, 213–222.

Schmelz, E.M., Roberts, P.C., Kustin, E.M., Lemonnier, L.A., Sullards, M.C., Dillehay, D.L., and Merrill, A.H. (2001). Modulation of Intracellular β-Catenin Localization and Intestinal Tumorigenesis in Vivo and in Vitro by Sphingolipids. Cancer Res. *61*, 6723–6729.

Sekiya, S., and Hafez, E.S.E. (1977). Physiomorphology of Twin Transfusion Syndrome: A Study of 86 Twin Gestations. Obstet. Gynecol. *50*, 288–292.

Shafer, B., Onishi, K., Lo, C., Colakoglu, G., and Zou, Y. (2011). Vangl2 Promotes Wnt/Planar Cell Polarity-like Signaling by Antagonizing Dvl1-Mediated Feedback Inhibition in Growth Cone Guidance. Dev. Cell *20*, 177–191.

Shaw, G.M., Todoroff, K., Velie, E.M., and Lammer, E.J. (1998). Maternal illness, including fever, and medication use as risk factors for neural tube defects. Teratology *57*, 1–7.

Shaw, G.M., Lu, W., Zhu, H., Yang, W., Briggs, F.B.S., Carmichael, S.L., Barcellos, L.F., Lammer, E.J., and Finnell, R.H. (2009a). 118 SNPs of folate-related genes and risks of spina bifida and conotruncal heart defects. BMC Med. Genet. *10*, 49.

Shaw, G.M., Carmichael, S.L., Vollset, S.E., Yang, W., Finnell, R.H., Blom, H., Midttun, Ø., and Ueland, P.M. (2009b). Mid-Pregnancy Cotinine and Risks of Orofacial Clefts and Neural Tube Defects. J. Pediatr. *154*, 17–19.

Shoob, H.D., Sargent, R.G., Thompson, S.J., Best, R.J., Drane, J.W., and Tocharoen, A. (2001). Dietary methionine is involved in the etiology of neural tube defect-affected pregnancies in humans. J. Nutr. *131*, 2653–2658.

Stockmann-Juvala, H., and Savolainen, K. (2008). A review of the toxic effects and mechanisms of action of fumonisin B1. Hum. Exp. Toxicol. *27*, 799–809.

Teraoka, S.N., Telatar, M., Becker-Catania, S., Liang, T., Önengüt, S., Tolun, A., Chessa, L., Sanal, Ö., Bernatowska, E., Gatti, R.A., et al. (1999). Splicing Defects in the Ataxia-Telangiectasia Gene, ATM: Underlying Mutations and Consequences. Am. J. Hum. Genet. *64*, 1617–1631.

Torres, O., Matute, J., Gelineau-van Waes, J., Maddox, J.R., Gregory, S.G., Ashley-Koch, A.E., Showker, J.L., Zitomer, N.C., Voss, K.A., and Riley, R.T. (2014). Urinary fumonisin B1 and estimated fumonisin intake in women from high- and low-exposure communities in Guatemala. Mol. Nutr. Food Res. *58*, 973–983.

Uddin, M., Koenen, K.C., Aiello, A.E., Wildman, D.E., de los Santos, R., and Galea, S. (2010). Epigenetic and inflammatory marker profiles associated with depression in a community-based epidemiologic sample. Psychol. Med. *41*, 997–1007.

Unoki, M., and Nakamura, Y. (2003). Methylation at CpG islands in intron 1 of EGR2 confers enhancer-like activity. FEBS Lett. *554*, 67–72.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A.M. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. *35*, W71–W74.

Velie, E.M., Block, G., Shaw, G.M., Samuels, S.J., Schaffer, D.M., and Kulldorff, M. (1999). Maternal Supplemental and Dietary Zinc Intake and the Occurrence of Neural Tube Defects in California. Am. J. Epidemiol. *150*, 605–616.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev. Cell *19*, 713–726.

Wallingford, J.B., and Harland, R.M. (2002). Neural tube closure requires Dishevelled-dependent convergent extension of the midline. Development *129*, 5815–5825.

Wallingford, J.B., Niswander, L.A., Shaw, G.M., and Finnell, R.H. (2013). The Continuing Challenge of Understanding, Preventing, and Treating Neural Tube Defects. Science *339*, 1222002.

Wang, G., and Bieberich, E. (2010). Prenatal alcohol exposure triggers ceramide-induced apoptosis in neural crest-derived tissues concurrent with defective cranial development. Cell Death Dis *1*, e46.

Wang, M., and Marín, A. (2006). Characterization and prediction of alternative splice sites. Gene *366*, 219–227.

Wang, D., Yan, L., Hu, Q., Sucheston, L.E., Higgins, M.J., Ambrosone, C.B., Johnson, C.S., Smiraglia, D.J., and Liu, S. (2012). IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinforma. Oxf. Engl. *28*, 729–730.

Wang, G., Krishnamurthy, K., and Bieberich, E. (2009). Regulation of primary cilia formation by ceramide. J. Lipid Res. *50*, 2103–2110.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

Watanabe, D., Suetake, I., Tada, T., and Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. Mech. Dev. *118*, 187–190.

Wysoker, A., Tibbetts, K., McCowan, M., Homer, N., and Fennell, T. (2009). Picard.

Xiong, Y., Zhang, L., Holloway, A.K., Wu, X., Su, L., and Kebebew, E. (2011). MiR-886-3p Regulates Cell Proliferation and Migration, and Is Dysregulated in Familial Non-Medullary Thyroid Cancer. PLoS ONE *6*, e24717.

Yalcin, B., Adams, D.J., Flint, J., and Keane, T.M. (2012). Next-generation sequencing of experimental mouse strains. Mamm. Genome *23*, 490–498.

Yang, J.-H., Li, J.-H., Jiang, S., Zhou, H., and Qu, L.-H. (2013). ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. Nucleic Acids Res. *41*, D177–187.

Ybot-Gonzalez, P., Cogram, P., Gerrelli, D., and Copp, A.J. (2002). Sonic hedgehog and the molecular regulation of mouse neural tube closure. Development *129*, 2507 –2517.

Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J. Comput. Biol. *11*, 377–394.

Yu, W., McDonnell, K., Taketo, M.M., and Bai, C.B. (2008). Wnt signaling determines ventral spinal cord cell fates in a time-dependent manner. Dev. Camb. Engl. *135*, 3687–3696.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Z., Chang, H., Li, Y., Zhang, T., Zou, J., Zheng, X., and Wu, J. (2010). MicroRNAs: Potential regulators involved in human anencephaly. Int. J. Biochem. Cell Biol. *42*, 37–374.

Zhao, B., Li, L., and Guan, K.-L. (2010). Hippo signaling at a glance. J. Cell Sci. *123*, 4001–4006.

Biography

Deidre Renee Krupp was born in Medford, Oregon on January 9, 1987. She graduated first-in-class valedictorian from South Medford High School in 2004 and went on to attend Southern Oregon University in Ashland, Oregon. In 2009 she earned Bachelor's of Science degrees in Biology with Honors and Chemistry with Honors, graduating *magna cum laude*. Deidre then moved to North Carolina to attain her doctorate through the Molecular Genetics and Microbiology program at Duke University, where she was awarded a Chancellor's Scholarship. During her time at Duke, Deidre published two articles in peer-reviewed journals, namely "Transcriptome profiling of genes involved in neural tube closure during human embryonic development using long serial analysis of gene expression (long-SAGE)" (2012, *Birth Defects Research Part A*, co-first author) and "The missing genetic risk in NTDs, can exome sequencing yield an insight?" (2014, *Birth Defects Research Part A*, first author). Her 2012 paper was a finalist for the James G. Wilson Publication Award of the Teratology Society. Deidre is also a Scholar of the Ronald E. McNair Postbaccalaureate Achievement Program, a lifetime member of Phi Kappa Phi, a member of Omicron Delta Kappa, and a member of Sigma Xi.